Rating rules / Puzzles. Ordering the rules

Goto page Previous 1, 2, 3 ... 16, 17, 18, 19, 20, 21 Next

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

View previous topic :: View next topic

Author	Message
denis_berthier	Posted: Sat Jul 04, 2009 10:08 pm Post subject: quote

Joined: 19 Jun 2007
Posts: 655
Location: Paris, France

THE NRCZT-RATING OF NRCZT-WHIPS AS A GUIDE FOR DEFINING THE RATING OF ANY CHAIN OR ANY PATTERN

In the second page of the present thread (http://www.sudoku.com/boards/viewtopic.php?t=5995&postdays=0&postorder=asc&start=23), I asked:

> **denis_berthier wrote:**
>
> how can we rate chains based on subsets, with different lengths and different maximal sizes for the subsets they use?

Remember first that the complexity of a resolution rule is in its condition pattern, not in its conclusion.
Remember also that the kind of complexity I'm considering is mainly complexity of resolution rules and that it is extended to complexity of puzzles in the SER style (= complexity of the most complex rule necessary to solve it).
The question then was mainly: how can the presence of subsets in a chain be taken into account in the definition of chain length? For ALS-chains, e.g., there are two kinds of parameters: chain length and size of each of the ALSs; how should they be combined?

After all the time and work on the subject of ratings, the answer now seems very clear to me: as the nrczt-whips can solve almost all the puzzles (there are fewer than 1 in a million minimal puzzles that it can't solve), the nrczt-rating constitutes an almost universal rating, to which all the other ratings can be compared *statistically*.
It has very good properties, the first three of which are not shared with the widely used SER:
- pure logic definition,

- implementation independence,
- full supersymmetry,
- very strong correlation (0.95) with log(# chains), which shows that it is statistically a logarithmic measure of complexity,
- strong correlation (0.895) with the well established SER for the human solvable puzzles, in spite of their a priori very different definitions.

It is therefore justified to take it as a reference for defining ratings based on different kinds of chains or patterns.

Consider first chains of ALS, AHS and A-Fish.
- define the length of such a chain to be the sum of the sizes of all its defining subsets, a single being considered as a subset of length 1 (see remarks below for precisions);
- define the LS-rating of a puzzle as the length of the longest such chain necessary to solve it.
Remarks:
- the "restricted commons" don't count in the subset sizes, they more or less play the roles of left-linking candidates,
- for the most classical complementarity reasons: NS(5) = HS(4), HS(5) = NS(4), NS(6) = HS(3) ...
- for supersymmetry reasons: for n = 1, 2, 3, 4, NS(n), HS(n) and SHS(n) (i.e. Fish(n)) are all counted as a NS(n).
[Edited twice to take Ronk's remarks into account.]
This rating is supersymmetric and consistent with the NRCZT-rating: whenever a chain can be viewed according to the two POV (which, according to the general subsumption results, is the case for almost all the chains of ALS, AHS and A-Fish), its length will be the same according to the two POV.
Although there is currently no program computing ratings of ALS-chains in a way consistent with this definiton, it is almost certain (due to the subsumption theorems) that this LS-rating is very strongly correlated with the nrczt-rating.

Now, consider the more general case of nrczt-whips(Subsets), as defined in the "fully supersymmetric chains" thread. In such chains, a naked, hidden and super-hidden (fish) subset (modulo the target and all the previous right-linking candidates) can be taken as a right-linking object, in lieu of a mere candidate.
Define the length of such a whip(Subsets) as the sum of all the sizes of the right-linking objects (Subsets) it contains (a single candidate being still considered as a subset of length 1).
Define the NRCZT(Subsets)-rating as the length of the maximal whip(Subsets) necessary to solve it.
Again, this rating is supersymmetric and consistent: any chain that can be viewed according to several POV (whether some parts are considered as subsets modulo the target and the previous rlc's or as mere nrczt-chains) will have the same ratings for all these POV.

We know that, most of the extremely rare puzzles that can't be solved with simple

whips (fewer than one in a million) can be solved with nrczt(Subsets)-whips. It is very likely that most of the puzzles that can be solved by mere whips will have the same ratings when they can be considered according to the two POVs. The nrczt(Subsets)-rating is thus almost an extension of the nrczt-rating.

Finally, this definition applies to Paul Isaacson's whips with ALS inserts and to Allan Barker's cover sets (nets, for which we take the sum of the sizes of the "sets", disregarding the sizes of the "linksets"). It therefore allows comparisons of the complexities of the solutions obtained with the corresponding patterns.

Last edited by denis_berthier on Mon Jul 06, 2009 11:12 pm; edited 3 times in total

Back to top profile pm www

m_b_metcalf

Joined: 15 May 2006
Posts: 2177
Location: Berlin

Posted: Sun Jul 05, 2009 2:55 am Post subject: quote

> **eleven wrote:**
>
> @Mike: Can you tell me, what's the difference between your new generation algorithm and suexg (beside of that suexg starts with 0 clues, which i think does not matter) ?

AFAIK, *suexg* solves the empty grid and eliminates clues until the puzzle is minimal (someone wil have to correct me if I'm wrong). This is similar to my original method that generates a random grid and eliminates. The new method adds random clues to an empty grid, notionally from the candidate list. This can lead to an invalid puzzle (zero solutions), and that has to be handled. Also, as I have stated already, coming from below means you are potentially more likey to find a local minimum that is close to the absolute minimum than when coming from above.

> **Red Ed wrote:**
>
> @Mike: well now I'm confused because I thought this was a thread about rating puzzles and (for example) the influence of bias on those ratings; **not** about finding ways to introduce bias towards low numbers of clues.

Sorry for muddying the waters. I'll be more careful from now on.

> **Red Ed wrote:**
>
> ... then selected the top 10000 and bottom 10000 scores ...

Another biased and perhaps interesting sample is of those puzzles that have only 8 clue values. Or has that already been studied?

Regards,

Mike Metcalf

Back to top profile pm

ronk

Joined: 02 Nov 2005
Posts: 2385
Location: Southeastern USA

Posted: Sun Jul 05, 2009 3:49 am Post subject: quote

denis_berthier wrote:

> Notice that, for symmetry reasons, an AHS(5) is considered to be an ALS(4)...

Would you please give an example for the meaning of *symmetry* in your statement above? Do you perhaps mean *equivalence*?

Can we correctly assume that "*an ALS(5) is considered to be an AHS(4)*" too?

Would the "(5) --- (4)" relationships be "modulo" the quantity of fills (clues + placements) in the unit (row, column, box)?

Back to top profile pm

eleven

Joined: 10 Feb 2008
Posts: 319

Posted: Sun Jul 05, 2009 4:47 am Post subject: quote

m_b_metcalf wrote:

> AFAIK, *suexg* solves the empty grid and eliminates clues until the puzzle is minimal (someone wil have to correct me if I'm wrong).

No, as you can see from the comments in my link:

Code:

```
// add a random clue and solve it. No solution ==>
remove it again.
// Not yet a unique solution ==> continue adding clues
    if(m2<1)A[i1]=0;if(m2!=1)goto mr1;

//now we have a unique-solution sudoku. Now remove
clues to make it minimal
...
```

So i wondered, why your algorithm produced a list with a different clue distribution.

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 540

Posted: Sun Jul 05, 2009 4:59 am Post subject: quote

Mike, thanks for suggesting looking at 8-clue puzzles. One way to do that is to reverse the terms in the "balance" score that I suggested before, so the most significant digit of the score is the number of clue-values that do not appear at all.

Results below (column 1 = SER; 2 = nr top 10000 puzzles w. that SER; 3 = nr bottom 10000 puzzles w. that SER). The "top" puzzles are those with 8 clues only. The "top" puzzles are much more likely to produce SER 1.x puzzles than "bottom" puzzles are. The result at SER 6.6 is there again, too:

Code:

```
THEVALIDITY = valid_25
THESCORE    = score_cluecounts_rev
Got ratings: 1.200000 to 9.300000, tens=10
```

```
nv = 297838
12      337     122
15      2867    1782
17      153     106
20      1344    2131
23      74      481
25      102     110
26      471     578
28      140     173
30      71      129
32      21      16
34      53      79
36      12      16
38      6       2
40      3       12
42      389     490
44      50      77
45      245     209
46      35      15
47      1       0
50      1       4
52      1       0
56      120     80
57      11      11
62      3       3
65      14      3
66      1255    549
67      197     62
68      38      48
69      17      15
70      11      12
71      793     901
72      635     1063
73      149     273
74      7       14
75      9       6
76      46      19
77      32      6
78      38      7
79      16      1
80      3       0
82      9       6
83      101     176
84      67      118
85      19      36
86      0       4
88      3       7
89      14      25
90      14      22
91      3       0
92      0       1
top: mean rating = 3.865810
bot: mean rating = 3.975610
Log nr combinations: 12942.070285
Natural occurrences: 0/10000
```

If anything, this is a better demonstration than my earlier one. Good suggestion!

Now I need to change focus and do the same for two different generators. Mike, can you send me your puzzles?

Back to top profile pm

m_b_metcalf Posted: Sun Jul 05, 2009 5:20 am Post subject: quote

eleven wrote:

> So i wondered, why your algorithm produced a list with a different clue distribution.

Joined: 15 May 2006
Posts: 2177
Location: Berlin

Hmm, so do I. If I wanted to start a scientific investigation, I'd start with the random number generators. I use a commercial one that is described in the literature:

Quote:

> The RANDOM_NUMBER generator uses two separate congruential generators together to produce a period of approximately 10**18, and produces real pseudorandom results with a uniform distribution in (0,1). It accepts two integer seeds, the first of which is reduced to the range [1, 2147483562]. The second seed is reduced to the range [1, 2147483398]. This means that the generator effectively uses two 31-bit seeds.
>
> For more information on the algorithm, see the following:
>
> Communications of the ACM vol 31 num 6 June 1988, titled: Efficient and Portable Combined Random Number Generators by Pierre L'ecuyer. Springer-Verlag New York, N. Y. 2nd ed. 1987, titled: A Guide to Simulation by Bratley, P., Fox, B. L., and Schrage, L. E.

My program may, of course, have all sorts of other bugs, but it's built on that firm foundation.

Regards,
Mike Metcalf

Back to top

m_b_metcalf

Posted: Sun Jul 05, 2009 5:27 am Post subject:

Red Ed wrote:

> Mike, thanks for suggesting looking at 8-clue puzzles.
>
> Now I need to change focus and do the same for two different generators. Mike, can you send me your puzzles?

Joined: 15 May 2006
Posts: 2177
Location: Berlin

You're welcome, and certainly. I have the PM with your e-mail address.

So far I have produced files from:

1) my old generator with weak removal algorithm (not interesting)

2) my old generator with improved removal

3) my old generator working from 1 grid

4) my old generator with biased removal

5) my new generator

6) (in prepartion) my old generator with all 9s removed from the solution grids.

Which do you want?

Regards,

Mike

Back to top profile pm

Red Ed

Posted: Sun Jul 05, 2009 5:40 am Post subject: quote

Joined: 06 Jun 2005
Posts: 540

Pick any two, please. And just enough grids from each that I should expect to be able to SE-rate the whole lot in less than 8 hours.

Ta muchly.

EDIT: oh my, Roddick's just taken the first set! :-o
EDIT2: what the ...?! One set apiece!

Back to top profile pm

Red Ed

Posted: Sun Jul 05, 2009 6:59 am Post subject: quote

Joined: 06 Jun 2005
Posts: 540

On Windows XP:

Code:

```
C:\foo> java -cp SudokuExplainer.jar
diuf.sudoku.test.Tester sudoku.txt result.txt
Exception in thread "main"
java.lang.NoClassDefFoundError: diuf/sudoku/test/Tester

C:\foo>
```

Anyone know how to rate multiple puzzles using SE on XP?

Back to top profile pm

m_b_metcalf

Posted: Sun Jul 05, 2009 7:19 am Post subject: quote

Joined: 15 May 2006
Posts: 2177
Location: Berlin

Red Ed wrote:

> On Windows XP:
>
> **Code:**
>
> ```
> C:\foo> java -cp SudokuExplainer.jar
> diuf.sudoku.test.Tester sudoku.txt result.txt
> Exception in thread "main"
> java.lang.NoClassDefFoundError:
> diuf/sudoku/test/Tester
>
> C:\foo>
> ```
>
> 

> Anyone know how to rate multiple puzzles using SE on XP?

You need to be in the appropriate directory. This .bat works for me on Vista and XP:

Code:

```
cd c:\sudoku
java -cp SudokuExplainer.jar diuf.sudoku.test.Tester
s.txt r.txt
pause
```

Back to top profile pm

Red Ed

Posted: Sun Jul 05, 2009 7:30 am Post subject: quote

Joined: 06 Jun 2005
Posts: 540

Bingo. Tvm.

So, before I burn far too many CPU cycles: do you happen to have any puzzle collections (from your own generator) with associated SE ratings? Or must I compute the SER myself?

Back to top profile pm

denis_berthier

Posted: Sun Jul 05, 2009 7:33 am Post subject: quote

Joined: 19 Jun 2007
Posts: 655
Location: Paris, France

ronk wrote:

> **denis_berthier wrote:**
>
> > Notice that, for symmetry reasons, an AHS(5) is considered to be an ALS(4)...
>
> Would you please give an example for the meaning of *symmetry* in your statement above?

I meant r <-> n or c <-> n or b <-> n super-symmetry.

ronk wrote:

> Can we correctly assume that "*an ALS(5) is considered to be an AHS(4)*" too?

Yes. We have to. Otherwise, we don't get the shorter possible length and we loose super-symmetry.

ronk wrote:

> Would the "(5) --- (4)" relationships be "modulo" the quantity of fills (clues + placements) in the unit (row, column, box)?

The "(5) --- (4)" relationships are the standard complementarity relationships. For me an ALS, AHS or A-Fish in a chain is just a LS, HS or SHS (Fish) modulo the restricted commons.

The next step, the generalisation of ALS-chains to whips(Subsets), is just generalising the almosting principle to the zt-ing principle.

Back to top profile pm www

m_b_metcalf

Posted: Sun Jul 05, 2009 7:40 am Post subject: quote

Joined: 15 May 2006
Posts: 2177
Location: Berlin

Red Ed wrote:

> Bingo. Tvm.
>
> So, before I burn far too many CPU cycles: do you happen to have any puzzle collections (from your own generator) with associated SE ratings? Or must I compute the SER myself?

No, but Denis must have for all the fist five sets. He's back tomorrow.

Regards,

Mike Metcalf

Back to top profile pm

denis_berthier

Posted: Sun Jul 05, 2009 7:43 am Post subject: quote

Joined: 19 Jun 2007
Posts: 655
Location: Paris, France

m_b_metcalf wrote:

> **eleven wrote:**
>
> > So i wondered, why your algorithm produced a list with a different clue distribution.
>
> Hmm, so do I. If I wanted to start a scientific investigation, I'd start with the random number generators.

I don't think the RNG can make much difference. I checked that the 1,000,000 puzzles in sudogen0_1M are different and uncorrelated.
Subtle differences in the way clues are added and then deleted may induce larger differences in the final output.
I think we need a precise description of the workings of the generators, in natural language. Reading the almost undocumented C code of suexg is harder for me than reading Chinese.

Back to top profile pm www

denis_berthier

Posted: Sun Jul 05, 2009 7:44 am Post subject: quote

Joined: 19 Jun 2007
Posts: 655
Location: Paris, France

m_b_metcalf wrote:

> **Red Ed wrote:**
>
> > Bingo. Tvm.
> >
> > So, before I burn far too many CPU cycles: do you happen to have any puzzle collections (from your own generator) with associated SE ratings? Or must I compute the SER myself?

> No, but Denis must have for all the fist five sets. He's back tomorrow.

Indeed, I'm back now. If you PM me a private email I'll send you the SER for the 5 sets.

Just received your PM. I'm sending them.

Back to top profile pm www

Display posts from previous: All Posts Oldest First 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