

The hardest sudokoku (new thread)

Re: The hardest sudokoku (new thread)
Dty champagne - Sat Oct 20, 2012 6:57 pm

A new update of the data base of potential hardest is available
The link is unchanged

[here](#)

The data base has now 56 448 puzzles, but nothing has been done on the new puzzles as far as exotic properties are concerned.

About half of the new puzzles have 25 clues and the distribution is the following for highest ratings

CODE: SELECT ALL										
ER	20	21	22	23	24	25				
11.9	0	1	4	2						
11.8	1	14	13	6						
11.7	1	25	18	7	4					
11.6	9	31	27	29	7					
11.5	42	95	103	41	2					
11.4	5	121	193	243	110	12				
11.3	41	579	829	1474	928	325				
11.2	72	766	1947	2438	1891	322				
11.1	59	714	1453	5421	3971	1141				
18.5	2284	3630	9723	6952	1572					

It's still too early to conclude for 24/25 clause, but it seems that highest cascade ratios will come for 21, 22 clause.

It's still to early to conclude for 24/25 clues, but it seems that highest serate ratings will come for 21_23 clues

champagne

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

Re: The hardest sudokoku (new thread)

Dty dobrichev - Sat Oct 20, 2012 7:46 pm

Congratulations for the new findings.

Is it easy for you to provide similar statistics for the 10+ puzzles from your cache?

QUOTE

dobrichev

2016 Supporter

Posts: 1831

Joined: 24 May 2010

Re: The hardest sudokoku (new thread)

Dty champagne - Sat Oct 20, 2012 8:11 pm

QUOTE

dobrichev wrote:

Congratulations for the new findings.

Is it easy for you to provide similar statistics for the 10+ puzzles from your cache?

Unhappily not because I don't keep all the ratings not in the data base.

The data base of potential hardest has ratings

diamonds 10.5 10.6
pearls 10.7 to 10.9
all ratings above 10.9

but I did not rate all puzzles that could be in that range. serate is too slow. The statistic can be given for the entire data base up loaded

One possibility to get such statistics is to re rate the data base of seeds using skfr. Part of that data base (about 5 millions puzzles) has been loaded in the Google project skfr (and downloaded 82 times)

My up to date data base of seeds has more than 7 millions puzzles of 20-24 clues and about 400 000 puzzles of 25 clues rating 10 and more.

Most of my generation is currently done in the 25 clues field

QUOTE

dobrichev

2016 Supporter

Posts: 1831

Joined: 24 May 2010

Re: The hardest sudokoku (new thread)

Dty JPF - Sat Oct 27, 2012 2:35 pm

Here is the link of an article recently published by Nature:

[The Chaos Within Sudoku](#)

JPF

QUOTE

JPF

2017 Supporter

Posts: 1847

Joined: 06 December 2005

Location: Paris, France

Re: The hardest sudokoku (new thread)

Dty champagne - Sat Oct 27, 2012 7:31 pm

Hi JPF,

Interesting article although my maths background is not adequate for a full understanding.

In that rating, Platinum Blonde has the top level, passing Golden Nugget.

Applying the exocet and the "abi loop", that puzzle (PB) collapse quickly.

Golden Nugget has a lower rating but a higher resistance on our side, even if Leren has a good solution.

The context is different, so comparing both approaches is not so important.

It would be interesting however, if you have the contact, to know how would be rated in that scale our entire file of "potential hardest" (based on serate ratings).

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

Re: The hardest sudokoku (new thread)

Dty JPF - Mon Oct 29, 2012 3:53 am

In June 2012 I posted that

QUOTE

champagne wrote:

the file of potential hardest puzzles has shown it had a cut-off to high for the search done on exotic patterns used to solve such puzzles.
I prepared in the skfr google project the entire file of puzzles rating (skfr) 10.0 and more I had available
warning: that file contains all the puzzles I used as seed in my search including already known (see the readme file)
[5million puzzles file](#)

That file has been downloaded 82 times, which surprised me.
I discarded it for several reason, one is that the file is now obsolete, the second that likely all current potential users had already downloaded it.

I have currently 7.5 millions such puzzles (nearly 500k with 25 clues).

The files looks now too big to be loaded in once. I intend to load now one file per number of clues.
May be users who downloaded the previous file have others ideas or comments

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

10.0+ collection

Dty dobrichev - Mon Oct 29, 2012 11:15 am

It is essentially a collection of 10+ puzzles rated by slightly different algorithm (skfr).
Since the target is not the number of clues (like for 17s or 39s) I see no reason to split it by number of clues. If there are technical reasons, it is better to split it by difficulty.
If there is interest maybe in the future it should be updated, including adding puzzles by other sources, and maybe simplifying (omitting) the credits.

I know the determinant reason in defining the threshold of your "main" hardest collection was the collection size. Do you have some expectations how 10+ would grow in time?

QUOTE

dobrichev

2016 Supporter

Posts: 1831

Joined: 24 May 2010

Re: The hardest sudokoku (new thread)

Dty champagne - Mon Oct 29, 2012 1:50 pm

Hi mladen ,

some disordered comments to your post.

A) The file I published is my data base of seeds used in the search of "potential high ratings"
It does not contains the "credit" if you mean the reference to the first known publication
The rating has been lost as well.
the puzzle is shown in it's maxdet form (always "9" in first cell).
I know that the rating is 10+ or more just because overtime I used a cut off of 10.2 at the start, and 10.0 later (skfr definiton) to extract puzzles to use in the next cycle.

B)Splitting the file is just in my mind a practical measure. I assume that in most cases, users will rebuild the entire file.
I had to split the data base to avoid problems with ACCESS: I have now a separated data base for 25 clues.
The number of clues is the only parameter easy to compute and stable in my process. (in the process a "n" clues seed generates a "n" clues puzzle)
In my view, the process is close to the end for 20_22 clues. Having a split per clue would give a way to check it.

C)How the number of 10+ will grow is unpredictable.
I am basically looking for 10.5/10.6 diamonds 10.7/10.9 pearls and 11.x puzzles where should be the hardest.
The last external contribution to my data base is from you (if I except the pattern game). It's long long ago
I have not heard of anybody else working currently on that topic.

D)Where am I
I am acting in 2 directions.
Continuing exploration of the 20_24 clues field but eleven made such a good job that most of the new puzzles are on the low side of the rating range.
Starting exploration of the 25 clues field. The start is very slow, but my experience of the pattern game is that you must not be discouraged.
When eleven started a deep exploration of the 20_23 clues area, many puzzles had been produced giving a sound base of seeds.
In the 25 clues field, the base of seeds has to be created and this is always a very slow process. (At least I know no quick way to do it)
As I told earlier, thinking that the peak value for the ratings has been reached, I lowered the cut off to 10
I am still using serate to rate the "potential hardest". It is nearly half of the runtime.
Any free cycles could help.

To conclude, as it is, the file is just a sample file of high ratings available for studies in that field. To split it in the way you suggest, it would be necessary to re rate it (which would be possible using skfr knowing that the highest rating are in a separate file with serate's rating).

Last edited by champagne on Tue Oct 30, 2012 8:15 am, edited 3 times in total.

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

Re: The hardest sudokoku (new thread)

Dty dobrichev - Mon Oct 29, 2012 2:58 pm

QUOTE

champagne wrote:

The file I published is my data base of seeds used in the search of "potential high ratings"
It does not contains the "credit" if you mean the reference to the first known publication

That means most of the puzzles are not published previously, collection is initially composed by you, and therefore is credited to you. Due to the all possible complications I proposed not to credit individual puzzles.

QUOTE

champagne wrote:

The rating has been lost as well.

I can rate them with the latest skfr.

QUOTE

champagne wrote:

the puzzle is shown in it's maxlex form.

Hope you meant minlex, the dialect which is implemented in my and gsf's tool and is not at 100% precise row-minlex. If so - thank you.

QUOTE

champagne wrote:

The number of clues is the only parameter easy to compute and stable in my process. (In the process a "n" clues seed generates a "n" clues puzzle)

Then just do it in this way. Seems I misunderstood your question about others' ideas or comments.
I am not sure that anchoring the number of the givens is the best process. More in my next post.

QUOTE

champagne wrote:

How the number of 10+ will grow is unpredictable.

OK. Keep in mind that your Access limitations can play a trick on you.

Cheers,
MD

QUOTE

dobrichev

2016 Supporter

Posts: 1831

Joined: 24 May 2010

One pass trough hardest top 7

Dty dobrichev - Mon Oct 29, 2012 4:14 pm

The last night I did an experiment how many hard puzzles could be found after depth relabeling followed by minimization of the appeared non-minimals.
I use the term relabeling for pattern-preserving transformation, where the givens' positions are fixed, and (x+x) transformation is performed on a seed puzzle.

The source is the 7 topmost puzzles in the champagne's hardest list (of 56448 puzzles):

CODE: SELECT ALL

98.7.....6.....6.5.....4.....5.3.....79.....5.....2.....1.85.....9.....1.....3.2.....11.9011.80/GP/champagne dty:1;22
98.7.....6.....87.....5.4.....3.5.....65.....0.....2.....1.86.....5.....1.3.....4.....211.9011.9011.60/GP/kx8;1523;23
12.....3.....4.....2.....62.....1.5.....4.....2.....6.....7.....9.....3.....8.....6.....11.9011.80/elev;xexod Flushy;2;23
.....39.....1.5.....3.8.....8.9.....6.7.....2.....1.4.....9.8.....5.....2.....6.....4.....7.....11.9011.9011.30/tax;601den-Nugget;3;21
.....4.....8.....8.....68.....71.....2.....5.....9.....95.....4.....2.....1.....7.....2.....4.....6.....121.9011.9011.90/elev;5;7;22
12.....3.....4.....5.....6.....7.....2.....6.....1.....3.....453.....8.....9.....451.....9.....8.....2.....711.9011.9011.60/elev;1;9;22
.....3.....6.....1.....2.....7.....7.....4.....9.....6.....3.....4.....1.....7.....2.....3.....5.....5.....6.....98.....5.....11.9011.20;1.20/elev;2;8;22

Mutations and filtering:
1) Relabeling in depth 6. Multiple-solution puzzles are ignored. No check for minimality at this step. Canonicalize and remove duplicates. 4721475 puzzles.
2) Splitting minimals (3319139) and non-minimals (1402316), remove invalid puzzles due to unknown slw bug (20).
3) Rate minimals using skfr. These 35 11+ appear (which include the seed I hope). 35.3319139 = **1.99K** with seed counted, and 28.3319139 = **1.118K** with seed excluded.

Hidden Text: Show

4) Minimize non-minimals (using some still unstable process). Canonicalize. (1765173).
5) I forget to remove duplicates from minimized and already minimal puzzles, but it is a feature list. 21.1765173 = **1.84K**, better than above.

Hidden Text: Show

6) Done.
The whole exercise took 21 hours - that is the time stamp difference between latest and first files in the directory. This time include several hours of idle CPU and much time in single-thread load, although the relabeling and later the rating have been done in parallel.

I know most of the listed puzzles are already discovered but hope there are few new 11+ there. I have no plans to rate the possible new ones with SE.

Cheers,
MD

QUOTE

dobrichev

2016 Supporter

Posts: 1831

Joined: 24 May 2010

Re: One pass trough hardest top 7

Dty colini - Mon Oct 29, 2012 8:51 pm

QUOTE

dobrichev wrote:

.....I know most of the listed puzzles are already discovered but hope there are few new 11+ there....

these are new / not in champagne's hardest list

CODE: SELECT ALL

.....12.....3.....4.....4.2.5.....1.4.....5.6.....7.....8.....9.....7.....2.....7.....5.....9.....6.....3.....8.....
.....12.....3.....4.....5.4.6.....1.5.6.....3.7.....8.....9.....4.....2.....6.....7.....8.....9.....5.....
.....12.....3.....4.....4.1.5.4.....6.....5.....3.....7.....9.....3.....8.....6.....2.....5.....7.....5.....9.....8.....6.....
.....12.....3.....4.....4.5.6.....3.....7.....486.....9.....5.....8.....45.....9.....1.....228.6.....
.....12.....3.....4.....4.2.5.....2.....6.....4.....3.....7.....8.....9.....9.....5.....1.....9.....8.....7.....6.....
.....12.....3.....4.....4.2.5.....1.....6.....4.....3.....7.....8.....9.....8.....1.....5.....9.....8.....6.....7.....6.....
.....12.....3.....4.....5.....6.....7.....3.....7.....486.....9.....5.....48.....9.....26.....8.....7.....2.....
.....12.....3.....4.....5.....6.....1.....7.....8.....6.....9.....54.....49.....8.....1.....22.6.....9.....
.....12.....3.....4.....5.....6.....1.....7.....8.....6.....4.....9.....54.....58.....8.....1.....22.6.....9.....
.....12.....3.....4.....5.....6.....2.....6.....3.....7.....4.....9.....4.....2.....6.....7.....5.....9.....8.....5.....
.....12.....3.....4.....5.....6.....3.....2.....457.....8.....6.....7.....45.....8.....1.....996.....7.....
.....12.....3.....4.....5.....6.....1.....2.....457.....8.....6.....7.....45.....8.....3.....996.....7.....
.....12.....3.....4.....4.2.5.....5.....6.....4.....7.....4.....8.....9.....6.....5.....3.....6.....9.....7.....6.....
.....12.....3.....4.....4.2.5.....2.....6.....4.....3.....7.....8.....9.....8.....6.....5.....3.....7.....8.....6.....9.....6.....
.....12.....3.....4.....4.2.5.....6.....7.....1.....6.....5.....8.....9.....5.....7.....1.....4.....9.....8.....3.....6.....
.....12.....3.....4.....2.....5.....6.....4.....1.....6.....7.....5.....8.....9.....5.....4.....6.....9.....7.....5.....3.....8.....
.....12.....3.....4.....5.....6.....1.....7.....458.....9.....6.....8.....85.....9.....24.....8.....3.....2.....
.....12.....3.....4.....5.....6.....1.....7.....458.....9.....6.....8.....85.....9.....2.....334.....8.....
.....12.....3.....4.....5.....6.....1.....7.....458.....9.....6.....8.....58.....9.....2.....334.....8.....

Perhaps this shows that the 6-depth search was more than the depth used by eleven

Maybe a 4 or 5 depth search on more of the puzzles would reveal even more. ?

I havent got round to using skfr yet - maybe an idiots guide is available !!! 😊

C

QUOTE

colini

2016 Supporter

Posts: 2225

Joined: 05 May 2005

Location: Devon

Re: The hardest sudokoku (new thread)

Dty rank - Mon Oct 29, 2012 10:54 pm

QUOTE

dobrichev wrote:

the puzzle is shown in it's maxlex form.

Hope you meant minlex, the dialect which is implemented in my and gsf's tool and is not at 100% precise row-minlex.

This looks to be what champagne has called maxlex in previous posts. If so, and if it is to be renamed, it should not be renamed maxlex. As I understand it ...

minlex is the row-ordered minimum lexicographical-order of the solution

maxlex is the row-ordered maximum lexicographical-order of the clues

Staying with the maxtext term helps distinguish between the use of the solution grid in the first instance, and use of clues in the second instance. Moreover, it leaves open the use of the maxdex term for a different canonicalization based on the solution, should the need arise. The maxclex term, instead of maxtext, might be even better.

Re: The hardest sudokoku (new thread)

Dty champagne - Tue Oct 30, 2012 2:26 am

some quick comments to the last 3 posts

1) I edited my last post to use the term maxtext for the canonical form.

2) quick puzzle on mladen example
list 1: one puzzle is new for my data base, skfr rating ED=11.1/11.9.4
list 2: no new puzzle
more comments to come to-morrow on the process

3) quick result on colini's list
skfr ratings in the range 11_11.3 (45 seconds for the entire list)
14 new puzzles in the range 11_11.2

4)diot guide for skfr

if the file name is "this is my file", use the command

skfr -i"this is my file"
or
skfr -ithis is my file"

if the filename is "myfile" (no blank), just use the command

skfr -imylfe

if you use a multi-threads version, it is better to control the number of threads and to have 2 command lines

set OMP_NUM_THREADS=2
skfr -imylfe

the command
skfr -imylfe
should deliver 2 output files
myfile_rat.txt with the results
myfile_N_rat with errors in the rating process (should be empty)

4) complete guide
a readme file is in the zip files prepared by mladen.
I went trough that readme file, we have to improve it.

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

Re: One pass trough hardest top 7

Dty champagne - Tue Oct 30, 2012 9:52 am

QUOTE

dobrichev wrote:

The last night I did an experiment how many hard puzzles could be found after depth relabeling followed by minimization of the appeared non-minimals.
I use the term relabeling for pattern-preserving transformation, where the givens' positions are fixed, and (x+x) transformation is performed on a seed puzzle.
.....
The whole exercise took 21 hours - that is the time stamp difference between latest and first files in the directory. This time include several hours of idle CPU and much time in single-thread load, although the relabeling and later the rating have been done in parallel.
.....
I know most of the listed puzzles are already discovered but hope there are few new 11+ there. I have no plans to rate the possible new ones with SE.

Cheers,
MD

QUOTE

champagne

2017 Supporter

Posts: 7269

Joined: 02 August 2007

Location: France Brittany

The result, if we look at new entries, looks disappointing. Only one new puzzle rating 11.1 (skfr).
In fact, the test has been done in the area 21_23 clues. That area has been deeply searched during the last five years and my feeling is that more than 90% of possibles are in the data base.

Now some personal comments on the process you described:

The runtime grows very fast with the depth.
Within the pattern (relabeling for you), exceeding a depth +3 must be done with a specific target. **"in"**
Changing the pattern +1 is ok, +2 already very long +3 nearly infeasible. **"out"**

But the acceptable processes (+-3in +1 out) do not extend easily the area where puzzles are searched.
Finding new seeds is a key point and the process you describe is one way to find new seeds in virgin areas.

In the pattern game, where the pattern is locked, I tested both a guided scan and a deep search.
The guided scan has better chances to give seeds covering all the field, but it is very expensive. Anyway, it is not working here.
Looking for fresh seeds, you can expand any puzzle, but you have then to have a highly permissive filter for the next cycle(s).

I am running a +4 in search in the 25 clues area. The yield is around 3% of fresh seeds with my cut_off at the skfr rating 10.0.
I process about 1000 puzzles per core/day, very very low. I get less fresh seeds than the number of puzzles processed, but I hope this will search in new areas in the next cycle.

In the pattern game, I often extend the search to +4 for highest ratings, usually not more.

In the current game, I did not find Patrice's 10.2 submission (I never use submissions for the generation, it would pollute my test). I compared the puzzle with my collection of ratings over 10. The closest puzzle seems to differ by 9 clues. A +9 would not be acceptable in the scope of the game. The results would come after game closure.

Last comment : I never keep non minimal puzzles.
In the pattern game, it has no value.
In the search for potential hardest, as I am studying "n" clues when "n-1" is nearly covered, it has a poor added value.

QUOTE

dobrichev wrote:

the puzzle is shown in it's maxlex form.

That means most of the puzzles are

