SAT Based approach to solving sudoku

THIRD YEAR PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BSC. IN COMPUTATIONAL MATHEMATICS

BY

Ashwot Acharya, Bishesh Bohora, Supreme Chaudhary, Lakki Thapa

DEPARTMENT OF MATHEMATICS
KATHMANDU UNIVERSITY
DHULIKHEL, NEPAL

1 CNF Conversion for 4×4 sudoku

1.1 What is CNF?

A boolean formula is said to be in a conjective normal form if it is a conjuction of clauses , where each clause is a disjunction of literals.Huth and Ryan 2004

To show how a sudoku is converted to a Conjective Normal Form (CNF) let us consider a basic 4×4 sudoku like the following:

	2		
		3	
	3		
		1	

1.2 Basic representation

Each cell in a 4×4 sudoku has a total of 4 different possible values for the ease of representation let us consider the following :
i: row number
j: column number
k: possible input value
$i, j, k \in \{1, 2, 3, 4\}$
And hence any cell x can be represented as : x_{ijk}

1.3 Clause for each cells

Since there only exist 1 input in each cell it can be written in terms of logic as :

$$(x_{ij1} \vee x_{ij2} \vee x_{ij3} \vee x_{ij4})$$

so for first cell we can represent the logic as : $(x_{111} \vee x_{112} \vee x_{113} \vee x_{114})$
rewriting for each cell we get the following:

$\bigvee_{k=1}^{4} x_{11k}$	x_{122}	$\bigvee_{k=1}^{4} x_{13k}$	$\bigvee_{k=1}^{4} x_{14k}$
$\bigvee_{k=1}^{4} x_{21k}$	$\bigvee_{k=1}^{4} x_{22k}$	x_{233}	$\bigvee_{k=1}^{4} x_{11k}$
$\bigvee_{k=1}^{4} x_{32k}$	x_{323}	$\bigvee_{k=1}^{4} x_{33k}$	$\bigvee_{k=1}^{4} x_{34k}$
$\bigvee_{k=1}^{4} x_{41k}$	$\bigvee_{k=1}^{4} x_{42k}$	x_{431}	$\bigvee_{k=1}^{4} x_{44k}$

Now Since each cell has at most 1 digit we can represent in CNF form in the following way:

$$\neg(x_{111} \wedge x_{112}) \equiv (\neg x_{111} \vee \neg x_{112}) \quad \text{(by De Morgan's law)}$$

So Applying For each possible values in first row first column we get the following 6 clauses:

$$(\neg x_{111} \vee \neg x_{112})$$

$$(\neg x_{111} \vee \neg x_{113})$$

$$(\neg x_{111} \vee \neg x_{114})$$

$$(\neg x_{112} \vee \neg x_{113})$$

$$(\neg x_{112} \vee \neg x_{114})$$

$$(\neg x_{113} \vee \neg x_{114})$$

The above is repeatitive pattern for each cell of the sudoku for example in cell 2 we will get the following

$$(\neg x_{121} \vee \neg x_{122})$$

$$(\neg x_{121} \vee \neg x_{123})$$

$$(\neg x_{121} \vee \neg x_{124})$$

$$(\neg x_{122} \vee \neg x_{123})$$

$$(\neg x_{122} \vee \neg x_{124})$$

$$(\neg x_{123} \vee \neg x_{124})$$

and so forth for the remaining cells with total of $16 \times 6 = 96$ clauses

The values in row and column of sudoku are unique i.e non repeatitive so row and column can have 1 to 4 in a non repetitive pattern the same follows for the sub grid of the sudoku .

1.4 uniqueness of row and column

1.4.1 For row

One digit on row must always appear only one in that row , to represent that for uniqueness of digit "1" :

$$\neg x_{111} \vee \neg x_{121}$$

$$\neg x_{111} \vee \neg x_{131}$$

$$\neg x_{111} \vee \neg x_{141}$$

$$\neg x_{121} \vee \neg x_{131}$$

$$\neg x_{121} \vee \neg x_{141}$$

$$\neg x_{131} \vee \neg x_{141}$$

uniqueness of digit "2" :

$$\neg x_{112} \vee \neg x_{122}$$

$$\neg x_{112} \vee \neg x_{132}$$

$$\neg x_{112} \vee \neg x_{142}$$

$$\neg x_{122} \vee \neg x_{132}$$

$$\neg x_{122} \vee \neg x_{142}$$

$$\neg x_{132} \vee \neg x_{142}$$

Similarly follows for 3 and 4 and for all four rows

1.4.2 For Column

for uniqueness of digit "1" in the first column , for uniqueness of digit "1" :

$$\neg x_{111} \vee \neg x_{211}$$

$$\neg x_{111} \vee \neg x_{311}$$

$$\neg x_{111} \vee \neg x_{411}$$

$$\neg x_{211} \vee \neg x_{311}$$

$$\neg x_{211} \vee \neg x_{411}$$

$$\neg x_{311} \vee \neg x_{411}$$

uniqueness of digit "2" in the first column :

$$\neg x_{112} \vee \neg x_{212}$$

$$\neg x_{112} \vee \neg x_{312}$$

$$\neg x_{112} \vee \neg x_{412}$$

$$\neg x_{212} \vee \neg x_{312}$$

$$\neg x_{212} \vee \neg x_{412}$$

$$\neg x_{312} \vee \neg x_{412}$$

and the same goes for digits '3' and digits '4' and follows for the remaining column i.e for each column there is $6 \times 4 = 24$ statements, and for column as a whole there are 96 statements the same goes for no of clauses for rows.

1.4.3 for 3×3 sub grid

In each subgrid there can only be one unique digits i.e there is no repetition of the digits in the subgrid. So In terms of logic it can be written as :
for example no repetition of digit '1' in the first subgrid of 2

$$\neg x_{111} \vee \neg x_{112}$$

$$\neg x_{111} \vee \neg x_{211}$$

$$\neg x_{111} \vee \neg x_{221}$$

$$\neg x_{121} \vee \neg x_{211}$$

$$\neg x_{121} \vee \neg x_{221}$$

$$\neg x_{211} \vee \neg x_{221}$$

It follows similarly for all digits 2 to 4 in the similar pattern and for each of the subgrids.

2 Writing of the CNF as a whole

The entire length of the CNF of the 4 sudoku will be pretty long but is written by AND-ing of each of the above clauses to create a long boolean Satisfiaility problem .
A snippet of the CNF form as a whole may be as follows :
uniqueness of the digit in the first cell

$$(\neg x_{111} \vee \neg x_{112}) \wedge (\neg x_{111} \vee \neg x_{113}) \wedge (\neg x_{111} \vee \neg x_{114}) \wedge (\neg x_{112} \vee \neg x_{113}) \wedge (\neg x_{112} \vee \neg x_{114}) \wedge (\neg x_{113} \vee \neg x_{114}) \wedge$$

$$(\neg x_{121} \vee \neg x_{122}) \wedge (\neg x_{121} \vee \neg x_{123}) \wedge (\neg x_{121} \vee \neg x_{124}) \wedge (\neg x_{122} \vee \neg x_{123}) \wedge (\neg x_{122} \vee \neg x_{124}) \wedge (\neg x_{123} \vee \neg x_{124}) \wedge ...$$

References

Huth, Michael and Mark Ryan (2004). *Logic in Computer Science: Modelling and Reasoning about Systems*. Cambridge: Cambridge University Press. ISBN: 978-0-521-54310-1.