Sudoku is Hard

Eve Routledge

Durham University

NP complete

Reduction

karp reduction

Sudoku $\leq_p$ Latin Square $\leq_p$ Triangulate a Tripartite Graph $\leq_p$ 3-SAT $\leq_p$ SAT

Def^n: A valid **Sudoku** puzzle is a function $S : i, j \rightarrow x$ for values i,j $\in \{1, ..., D^2\}$ and $x \in \{0, ..., D^2\}$ satisfying the following:

- for all $a, b, c \in \{1, ..., D^2\}$ with $S(a, b) \neq 0$ and $S(a, c) \neq 0$, then $S(a, b) \neq$ S(a,c)
- for all $a, b, c \in \{1, ..., D^2\}$ with $S(a, b) \neq 0$ and$S(c, b) \neq 0$, then $S(a, b) \neq$ S(c,b)
- for all $a, b, c, d \in \{1, ..., D^2\}$ with a mod $D = c$ mod D, b mod $D = d$ mod D, $S(a, b) \neq 0$ and $S(c, d) \neq 0$, then $S(a, b) \neq S(a, c)$

It is completed if $x \in \{1, ..., D\}$.

Def^n: A valid **Latin Square** puzzle is a function $L : i, j \rightarrow x$ for values $i, j \in \{1, .., D\}$ and $x \in \{0, ..., D\}$ satisfying the following:

- for all $a, b, c \in \{1, ..., D\}$ with $L(a, b) \neq 0$ and $L(a, c) \neq 0$ then $L(a, b) \neq L(a, c)$
- for all $a, b, c \in \{1, ..., D\}$ with $L(a, b) \neq 0$ and $L(c, b) \neq 0$ then $L(a, b) \neq L(c, b)$

It is complete or solved if for all $i, j \in \{1, ..., D\}$, $L(i, j) \neq 0$.

Def^n: A graph $G = (V, E)$ is tripartite if a partition V_1, V_2, V_3 exists such that the vertices are split into three sets with no edges between vertices that belong to the same set, i.e for all $(v_i, v_j) \in E$ if $v_i \in V_i$ then $v_j \notin V_i$.

Def^n: A **Triangulation** T of a graph G is a way to divide all edges into disjoint subsets T_i, each forming a triangle ($T_i = \{(v_1, v_2), (v_2, v_3), (v_3, v_1)\}$).

Def^n: A **boolean expression** With a set of boolean variables B and a collection of clauses C, with at most 3 literals (a literal is any $b \in B$ or its negation $\bar{b}$) in each, does a valid truth assignment exist that satisfies all clauses in C?

Cook Levin Theorem

SAT is NP-complete.

Sudoku $\leq_p$ Latin Square

Lemma: Let S be a Sudoku problem with the following construction

$$S(i, j) = \begin{cases} 0 & \text{when } (i, j) \in S_l \\ ((i - 1 \bmod n)n + \lfloor i - 1/n \rfloor + j - 1) \bmod n^2 + 1 & \text{otherwise} \end{cases} \quad (1)$$

where $S_l = \{(i, j) | \lfloor i - 1/n \rfloor = 0$ and $(j \bmod n) = 1\}$. Then there exists an augmentation S' to complete the sudoku puzzle if and only if the square L such that $L(i, j/n) = (S'(i, j) - 1)/n + 1$ for all $(i, j) \in S_l$ is a Latin square.

1	2	3	4	5	6	7	8	9
4	5	6	7	8	9	1	2	3
7	8	9	1	2	3	4	5	6
2	3	4	5	6	7	8	9	1
5	6	7	8	9	1	2	3	4
8	9	1	2	3	4	5	6	7
3	4	5	6	7	8	9	1	2
6	7	8	9	1	2	3	4	5
9	1	2	3	4	5	6	7	8

S(i,j), n=3, blanks removed

	2	3		5	6		8	9
	5	6		8	9		2	3
	8	9		2	3		5	6
2	3	4	5	6	7	8	9	1
5	6	7	8	9	1	2	3	4
8	9	1	2	3	4	5	6	7
3	4	5	6	7	8	9	1	2
6	7	8	9	1	2	3	4	5
9	1	2	3	4	5	6	7	8

S(i,j)

1	4	7
4	7	1
7	1	4

$S_l(i, j)$

1	2	3
2	3	1
3	1	2

L(i,j)

Latin Square $\leq_p$ Triangulated Tripartite Graph

Observe completing a partial Latin square is equivalent to triangulating tripartite graph, we map the Latin square to this through the following: given tripartite graph G=(V,E) label vertices in V_1 with distinct lables $\{r_1, ...r_n\}$, label vertices in V_2 with distinct lables $\{c_1, ...c_n\}$ and label vertices in V_3 with distinct lables $\{e_1, ...e_n\}$. Add edges such that:

- If $L(i, j) = 0$ then add the edge (r_i, c_j)
- If for all $i \in [0, ..., n]$ and constant j, $L(i, j) \neq k$ then add the edge (r_i, e_k)
- If for all $j \in [0, ..., n]$ and constant i, $L(i, j) \neq k$ then add the edge (c_j, e_k)

This graph has a triangulation iff L(i,j) can be solved.

Let us show every uniform tripartite graph is the above formulation of a Latin square.

Defn: A Latin framework LF for tripartite graph G, size (r,s,t) is a r by s array with values [1,...,t]. With constraints:

- Each row/column contain each element only once.
- If $(r_i, c_j) \in E$ then LF(i,j)=empty else LF(i,j)= k $k \in [1, ..., t]$
- If $(r_i, e_k) \in E$ then for constant i $LF(i, j) \neq k$
- If $(c_i, e_k) \in E$ then for constant j $LF(i, j) \neq k$

If r=s=t then LF is a latin square which can be completed iff G has a triangle partition.

Lemma: For graph G=(V,E) with |V|=n, there's a Latin framework of size (n,n,2n).

Define LF an n by n array. For $(r_i, c_j) \in E$ $LF(i, j) = 0$ else $LF(i, j) = 1 + n + ((i + j) mod n)$. LF is a latin framework. □

Lemma: Latin framework L (r,s,t) for uniform tripartite graph G. R(k) = the number of times k appears in L plus half $|e_k|$. Whenever $R(k) \geq r + s - t$ for $1 \leq k \leq t$, L can be extended to (r,s+1,t) to give L' in which $R'(k) \geq r + s + 1 - t$ for all $1 \leq k \leq t$. Lemma: Latin framework of size (r,s,s) can be extended to a Latin framework size (s,s,s).

Given a tripartite graph G, if it is not uniform then no triangulation exists, else we apply above to produce a latin framework of size (2n,2n,2n) in polynomial time. This is a Latin square which can be completed iff G has a triangulation. The latin square problem has been reduced to the triangulating a tripartite graph problem. □

Triangulated Tripartite Graph $\leq_p$ 3 SAT

3 SAT $\leq_p$ SAT

Given a truth assignment t check each clause is satisfied, if all are satisfied return True else False,

Stochastic Methods

References