Analysis of One-to-One Matching Mechanisms via SAT Solving: Impossibilities for Universal Axioms

Ulle Endriss

Institute for Logic, Language and Computation
University of Amsterdam

What is this about? The paper develops an approach to largely automate the generation of human-readable proofs for impossibility theorems in the field of matching. This approach relies on the power of modern SAT solving technology.

What is matching? Belongs to both game theory and social choice theory. Concerned with the design of mechanisms to match agents from two groups, based on their preferences. Think of job seekers and companies.

What is SAT solving? SAT is the NP-hard problem of deciding whether a set of clauses is satisfiable. Thanks to recent progress in AI and OR, modern solvers often can handle millions of clauses in a matter of seconds.

You want to design a matching mechanism for $n+n$ agents that is top-stable (mutual favourites are assigned to each other) and two-way strategyproof (nobody can benefit from misrepresenting their preferences). You don't manage. *Maybe it's just impossible?*

To *prove* it's impossible, use this approach . . .

Express your axioms in the formal language introduced in the paper:

top-stability

$$\forall_{\mathrm{P}} p.\forall_{\mathrm{N}} i.\forall_{\mathrm{N}} j.[\,(top^{\mathrm{L}}_{p,i} = j \wedge top^{\mathrm{R}}_{p,j} = i) \;\rightarrow\; p \rhd (i,j)\,]$$

strategyproofness for agents on the left and on the right

$$\forall_{\mathrm{P}} p.\forall_{\mathrm{P}} p'.\forall_{\mathrm{N}} i.\forall_{\mathrm{N}} j.\forall_{\mathrm{N}} j'.[(j \succ^{\mathrm{L}}_{p,i} j' \wedge p \sim^{\mathrm{L}}_i p') \rightarrow \neg(p \rhd (i,j') \wedge p' \rhd (i,j))]$$

$$\forall_{\mathrm{P}} p.\forall_{\mathrm{P}} p'.\forall_{\mathrm{N}} j.\forall_{\mathrm{N}} i.\forall_{\mathrm{N}} i'.[(i \succ^{\mathrm{R}}_{p,j} i' \wedge p \sim^{\mathrm{R}}_j p') \rightarrow \neg(p \rhd (i',j) \wedge p' \rhd (i,j))]$$

Having expressed your axioms this way, you see that they are universal.

Preservation Theorem: *If there exists a top-stable mechanism for $n+n$ agents that satisfies a given set of universal axioms, then also for $(n-1)+(n-1)$ agents.*

You are done if you can prove your conjecture for the special case of $3+3$ agents (it's false for $n \leq 2$). Encode this case in propositional logic using variables $x_{p\rhd(i,j)}$. For example, the first part of strategyproofness becomes:

$$\bigwedge_{i\in[3]} \bigwedge_{p\in[3!^{3+3}]} \bigwedge_{\substack{p'\in[3!^{3+3}] \\ \text{s.t. } p \sim^{\mathrm{L}}_i p'}} \bigwedge_{j\in[3]} \bigwedge_{\substack{j'\in[3] \\ \text{s.t. } j \succ^{\mathrm{L}}_{p,i} j'}} (\neg x_{p\rhd(i,j')} \vee \neg x_{p'\rhd(i,j)})$$

For your particular problem, you end up with a set of 4,805,568 clauses.

SAT solvers such as PicoSAT can analyse this set in around 1 second.

You find that the set you built is unsatisfiable. So now you know:

Impossibility Theorem: *For no $n \geq 3$ does there exist a matching mechanism that is both top-stable and two-way strategyproof.*

This is a strong variant of a seminal result due to Alvin Roth (1982).

You can use further tools to extract a minimal unsatisfiable subset, leading to a simple human-readable proof.

Paper and code available at tinyurl.com/satmatching.