Modelling and Solving the Stable Marriage problem using Constraint Programming

SM Formalisation

- Set ***n*** men $S_M = \{m_1, m_2, \ldots, m_n\}$
- Set ***n*** women $S_W = \{w_1, w_2, \ldots, w_n\}$

- Each man ranks the women in S_W in strict order of preference.
- Each woman ranks the men in S_M in strict order of preference.

- A matching ***M*** is a bijection between the men and women.
- We say a (man, woman) pair (m,w) **blocks *M*** if:
  - ***m*** prefers ***w*** to his partner in ***M***, and
  - ***w*** prefers ***m*** to her partner in ***M.***
- A matching that admits no blocking pair is said to be **stable**
  - Can't improve by making an arrangement outside the matching.
- SM was first formalised by David Gale and Lloyd Shapley in 1962.

Example Stable Marriage Instance

1:	2	4	1	3	1:	2	1	4	3
2:	3	1	4	2	2:	4	3	1	2
3:	2	3	1	4	3:	1	4	3	2
4:	4	1	3	2	4:	2	1	4	3

Men's preferences　　Women's preferences

Example Stable Matching

1:	2	(4)	1	3	1:	2	1	(4)	3
2:	(3)	1	4	2	2:	4	(3)	1	2
3:	(2)	3	1	4	3:	1	4	3	(2)
4:	4	(1)	3	2	4:	2	(1)	4	3

Men's preferences   Women's preferences

$$M = \{(1\,,4)\,,(2\,,3)\,,(3\,,2)\,,(4\,,1)\}$$

Extended Gale-Shapley Algorithm

assign each person to be free;
while (some man ***m*** is free)
    w := first women on ***m***'s list;
    if (***w*** is currently assigned to some man ***p***)
      assign ***p*** to be free;
    end if
    assign ***m*** to ***w***;
    foreach ($m' \in successors_w(m)$)
      delete (***m'***, ***w***);
    end loop
end loop

Algorithm complexity $O(n^2)$ – so linear in size of problem instance.

SM and CSP's

- In the last few years SM and CSP's have been the focus of much attention in the literature.
  - Gent et al: CP '01 with two SM encodings
  - Lustig and Puget, 2001
  - Green and Cohen, CP '03
  - Aldershof and Carducci, 1999

Motivation

- CP '01 paper by Gent et al. presented two ways to encode an instance of SM as a Constraint Satisfaction Problem (CSP).
  - First encoding: time taken to establish AC is ***$O(n^4)$*** - the variables domains after AC corresponded to the GS-lists.
  - Second encoding: time taken to establish AC is ***$O(n^2)$*** - the variables domains after AC corresponded to a weaker structure.
- Can we find an encoding that finds GS-lists in ***$O(n^2)$*** time?
- Shows that CP algorithms give rise to the same structure and the same time complexity as conventional methods.
- Many stable matching problems are NP-hard – can CP help us here?

Structure of SM

- The Gale-Shapley (GS) algorithm has two possible orientations:
  - **man-oriented** GS (MEGS) algorithm : where the men propose to the women.
  - **women-oriented** GS (WEGS) algorithm : where the women propose to the men.

- Optimality properties of each algorithm:
  - MEGS algorithm - **man-optimal** stable matching M_0 - simultaneously the best possible stable matching for all men.
  - WEGS algorithm - **woman-optimal** stable matching M_z- simultaneously the best possible stable matching for all women.

- Deletions that occur during an execution of either algorithm result in a set of reduced preference lists on termination of each algorithm:
  - The **MGS-lists** for the MEGS algorithm.
  - The **WGS-lists** for the WEGS algorithm.

- The intersection of the MGS-lists and the WGS-lists is called the **GS-lists**.
- The GS-lists have many important structural properties.

New CSP Encodings

- We present two new CSP encodings for SM.
- First encoding (*n*-valued) is simple and easy to understand, presents a natural way to represent SM as CSP.
- Second encoding (4-valued) is more complex but addresses the problem of establishing AC and finding the GS-lists in ***$O(n^2)$***. This will only briefly be mentioned here.

n-valued Encoding

- Encode SM instance ***I*** with ***n*** men and ***n*** women as a CSP instance ***J*** with ***2n*** variables.
- For each man $m_i \in S_M$ we introduce a variable x_i in ***J***.
- For each woman $w_j \in S_W$ we introduce a variable y_j in ***J***.
- The initial domain for each variable is:

$$dom(x_i) = dom(y_j) = \{1, 2, \ldots, n\}.$$

- Constraints:

1. $x_i \geq p \Rightarrow y_j \leq q \ (1 \leq i \leq n, 1 \leq p \leq n)$
2. $y_j \geq q \Rightarrow x_i \leq p \ (1 \leq j \leq n, 1 \leq q \leq n)$
3. $y_j \neq q \Rightarrow x_i \neq p \ (1 \leq j \leq n, 1 \leq q \leq n)$
4. $x_i \neq p \Rightarrow y_j \neq q \ (1 \leq i \leq n, 1 \leq p \leq n)$

In Constraints 1 and 4, ***j*** is such that $rank(m_i, w_j) = p$ and also $rank(w_j, m_i) = q$.
In Constraints 2 and 3; ***i*** is such that $rank(w_j, m_i) = q$ and also $rank(m_i, w_j) = p$.

n-valued Properties and Structure

- We prove that after AC propagation the variables domains correspond to the GS-lists.
- AC propagation in encoding takes ***$O(n^3)$*** time.
- We also prove that we can enumerate all stable matchings in a failure free manner.

4-valued Encoding

- Far more complex, extends second encoding in CP '01 paper.
- Name arises from the fact that each variable's domain contains 4 values
  - ***1*** - is always in the domain, ensures domain never becomes empty.
  - ***0*** - corresponds to a proposal.
  - ***2*** - corresponds to a deletion during the MEGS algorithm.
  - ***3*** - corresponds to a deletion during the WEGS algorithm.

- Gives us the GS-lists after AC propagation.

- AC can be established in ***O(n²)*** time.

- Failure free enumeration of all stable matchings also holds.

Summary

- Presented two new encodings for SM:
  - Time complexities ***$O(n^3)$*** and ***$O(n^2)$***.
  - AC propagation finds the GS-lists.
- Can be extended to SM with incomplete lists.
- First encoding has been extended to Hospitals / Residents problem (HR)
  - Will be presented at workshop in CP '05.
- Extension to NP-hard variants
  - HR with couples
  - HR with ties under weak stability