Homogeneous and Heterogeneous Island Models for the Set Cover Problem

Andrea Mambrini[1] Dirk Sudholt[2] Xin Yao[1]

University of Birmingham, Birmingham, UK

University of Sheffield, Sheffield, UK

May 7, 2012

Set Cover Problem

- $S = \{s_1, \cdots, s_m\}$ a set containing m elements
- $C = \{C_1, \cdots, C_n\}$ a collection of n non-empty subsets of S such that $\bigcup_{i=1}^{n} C_i = S$.
- Each set C_i has a cost $c_i > 0$.
- $X = x_1 \cdots x_n$ is a *selection* of C. C_i is in the selection X iff $x_i = 1$.
- The optimal solution to the SetCover problem is an X such that $\bigcup_{i:x_i=1} C_i = S$ and $\sum_{i:x_i=1} c_i$ is minimum.

SetCover is NP-hard so we look for polynomial approximation algorithms.

Greedy algorithm

The algorithm

Starting from an empty selection the greedy algorithm selects at each iteration the set that has the highest ratio between the number of uncovered elements and the cost of the set.

It provides an approximation ratio of H_m (the m-th armonic number, $O(\log m)$) but:

- it is not possible to make it parallel
- it is a deterministic algorithm (the approximation provided on a specific instance can't be improved running the algorithm again).

SetCover as a multi-objective optimization problem

A two-objective function to minimize $f(x) = (u(x), \text{cost}(x))$

- $\text{cost}(x)$ is the cost of the selection
- $u(x)$ is the number of uncovered elements

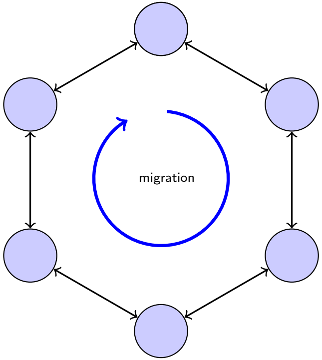
Algorithm 1 Global SEMO (GSEMO)

1. Choose an initial solution $s \in \{0,1\}$ uniformly at random
2. Initialize $P := \{s\}$
3. Repeat
   a) Choose $s \in P$ randomly
   b) Define s' by flipping each bit of s indipendently of the other bits with probability $1/n$
   c) Add s' to P
   d) Remove the dominated individuals from P

Theorem (Friedrich *et al.*, 2010)

For an empty initialisation and every SetCover *instance GSEMO finds an H_m-approximate solution in $O(m^2 n)$ generations.*

Island models

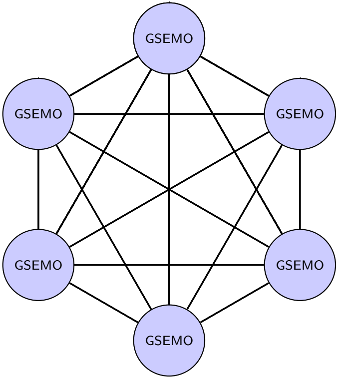


- Each island ideally runs on a different computational node
- Each island has its own population and runs an EA
- Individuals can migrate from an island to its neighbours

Measures

- Parallel runtime: the number of generations until at least one island has found a satisfactory solution
- Communication effort: the total number of individuals sent between islands, throughout a run of an island model

Homogeneous island model



- μ islands
- 2-objective fitness function (to minimize):

$$f(X) = (u(X), \mathrm{cost}(X))$$

The algorithm

For each island i:

- Simulate one generation of (global) SEMO
- Send a copy of the population $P_i^{(t)}$ to all neighboured islands with probability p.
- Unify $P_i^{(t)}$ with all populations received from other islands.
- Remove all dominated search points from $P_i^{(t)}$.

Analysis of the Homogeneous island model

Given a general result from Lässig and Sudholt (2011) on the analysis of Parallel EAs we get the following expected parallel runtime until an H_m-approximation is found:

- Parallel runtime: $O\left(\frac{m}{p} + \frac{nm^2}{\mu}\right)$

The expected communication effort is by a factor of $(m+1)pd\mu$ larger than the expected parallel time, where d is the degree of any node in the graph:

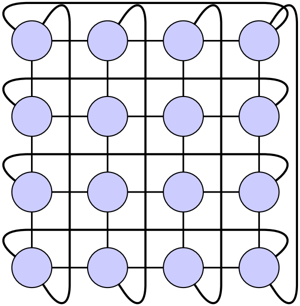
- Communication effort: $O\left(\mu^2 m^2 + p\mu nm^3\right)$

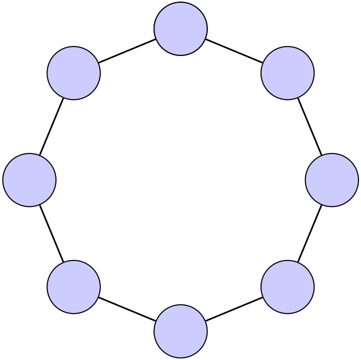
So for $\mu \leq pnm$:

- Parallel runtime: $O\left(\frac{mn^2}{\mu}\right)$
- Best parallel running time: $O\left(\frac{m}{p}\right)$ (for $\mu = pnm$)
- Communication effort: $O(p^2 n^2 m^4)$

Other topologies

Do we really need a complete topology?





Homogeneous island model - Results

Algorithm	parallel time bounds general b. $\rightsquigarrow$ best bound	comm. effort
Non-parallel (global) SEMO	$O(nm^2) \rightsquigarrow O(nm^2)$	0
(global) SEMO based homogeneous island models with topology...		
– complete ($\mu \leq pnm$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{m}{p}\right)$	$O(p^2n^2m^4)$
– grid ($\mu \leq (pnm)^{2/3}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/3}m^{4/3}}{p^{2/3}}\right)$	$O(pnm^3)$
– ring ($\mu \leq \sqrt{pnm}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/2}m^{3/2}}{p^{1/2}}\right)$	$O(pnm^3)$

- Trade-off between the parallel running time and the communication effort
- Less density $\rightarrow$ smaller highest amount of islands that still lead to linear speedup $\rightarrow$ higher parallel time best bound

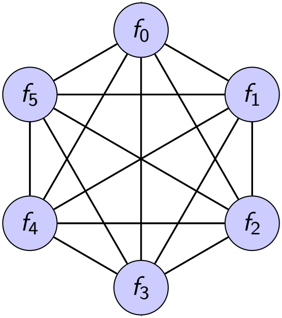
Issues of the homogeneous island model

- Each island has a population size up to m individuals:
  - A lot of individuals are exchanged during migration
  - Each island has to store up to $m\mu$ individuals after migration.
  - At each generation each island has to detect the non-dominated individuals among up to $m\mu$ individuals.
- Each island is performing the same task

Can we do better?

Can we come up with an island model that still provides linear speedup but in which each island perform different tasks, has a smaller population size and it is not necessary to perform non-dominated removal operations?

Heterogeneous island model



$c(X) = m - u(X)$ is the number of covered elements

- $m + 1$ islands
- each island runs a (1+1) EA (or RLS) on a different fitness function (to maximize)

$$f_i(X) = \begin{cases} nc_{\max} - \mathrm{cost}(X) & \text{if } c(X) = i \\ -|c(X) - i| & \text{if } c(X) \neq i \end{cases}$$

The algorithm

Each island i:

- Produce a global (or local) mutation $\tilde{X}_i^{(t)}$ of the individual $X_i^{(t)}$.
- Send a copy of $\tilde{X}_i^{(t)}$ to each other island.
- Choose $X_i^{(t+1)}$ with maximal f_i-value among $X_i^{(t)}$, $\tilde{X}_i^{(t)}$ and all immigrants.

Analysis of the heterogeneous island model

With an analysis similar to the one for GSEMO we get the following result:

Theorem

The heterogeneous island model with m islands finds an H_m-approximate solution for SETCOVER *in an expected parallel time of $O(nm)$. The expected communication effort is $O(nm^3)$.*

To fairly compare it with the homogeneous model we simulate the m islands on $\mu \leq m$ processors running up to $\lceil \frac{m+1}{\mu} \rceil$ islands on each processor. The parallel runtime is then slowed down to $O\left(\frac{nm^2}{\mu}\right)$.

Algorithm	parallel time bounds general b. $\rightsquigarrow$ best bound	comm. effort
Non-parallel (global) SEMO	$O(nm^2) \rightsquigarrow O(nm^2)$	0
(global) SEMO based homogeneous island models with topology. . .		
– complete ($\mu \le pnm$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{m}{p}\right)$	$O(p^2 n^2 m^4)$
– grid ($\mu \le (pnm)^{2/3}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/3} m^{4/3}}{p^{2/3}}\right)$	$O(pnm^3)$
– ring ($\mu \le \sqrt{pnm}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/2} m^{3/2}}{p^{1/2}}\right)$	$O(pnm^3)$
(1+1) EA (or RLS) based heterogeneous island models with migration policy. . .		
– complete ($\mu \le m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^3)$

Can we improve this result using a probabilistic migration policy?

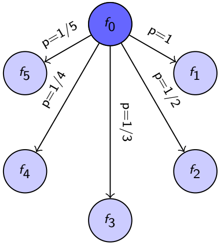
Migration policies – Uniform probabilistic

uniform probabilistic: each island sends migrants to every other island independently with a migration probability p.

Algorithm	parallel time bounds general b. $\rightsquigarrow$ best bound		comm. effort
(1+1) EA (or RLS) based heterogeneous island models with migration policy...			
– complete ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right)$	$\rightsquigarrow O(nm)$	$O(nm^3)$
– uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu p}\right)$	$\rightsquigarrow O\left(\frac{nm^2}{p}\right)$	$O(nm^3)$

Inter-island evolution is reduced $\rightarrow$ the reduced communication effort in a single generation is nullified by a larger parallel running time.

Migration policies – Non-Uniform Probabilistic



non-uniform probabilistic: each island i sends migrants to every other island $(i+k) \bmod (m+1)$ independently with probability $1/k$

Algorithm	parallel time bounds general b. ⇝ best bound		comm. effort
(1+1) EA (or RLS) based heterogeneous island models with migration policy. . .			
– complete ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right)$	⇝ $O(nm)$	$O(nm^3)$
– uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu p}\right)$	⇝ $O\left(\frac{nm^2}{p}\right)$	$O(nm^3)$
– non-uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right)$	⇝ $O(nm)$	$O(nm^2 \log m)$

Migration policies – Smart migration

$$f_i(X) = \begin{cases} nc_{\max} - \mathrm{cost}(X) & \text{if } c(X) = i \\ -|c(X) - i| & \text{if } c(X) \neq i \end{cases}$$

Do we really need to send individuals to islands where they would have negative fitness?

smart migration: each island i sends migrants to island $c(\tilde{X}_i)$, where $\tilde{X}_i$ is the offspring generated on island i.

Algorithm	parallel time bounds general b. $\rightsquigarrow$ best bound	comm. effort
(1+1) EA (or RLS) based heterogeneous island models with migration policy. . .		
– complete ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^3)$
– uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu p}\right) \rightsquigarrow O\left(\frac{nm^2}{p}\right)$	$O(nm^3)$
– non-uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^2 \log m)$
– smart migration ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^2)$

Results comparison

Algorithm	parallel time bounds general b. $\rightsquigarrow$ best bound	comm. effort
Non-parallel (global) SEMO	$O(nm^2) \rightsquigarrow O(nm^2)$	0
(global) SEMO based homogeneous island models with topology...		
– complete ($\mu \leq pnm$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{m}{p}\right)$	$O(p^2n^2m^4)$
– grid ($\mu \leq (pnm)^{2/3}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/3}m^{4/3}}{p^{2/3}}\right)$	$O(pnm^3)$
– ring ($\mu \leq \sqrt{pnm}$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O\left(\frac{n^{1/2}m^{3/2}}{p^{1/2}}\right)$	$O(pnm^3)$
(1+1) EA (or RLS) based heterogeneous island models with migration policy...		
– complete ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^3)$
– uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu p}\right) \rightsquigarrow O\left(\frac{nm^2}{p}\right)$	$O(nm^3)$
– non-uniform prob. ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^2 \log m)$
– smart migration ($\mu \leq m$)	$O\left(\frac{nm^2}{\mu}\right) \rightsquigarrow O(nm)$	$O(nm^2)$

Conclusion

Two parallel models that provably find H_m-approximated solutions to SetCover and provide linear speed-up compared to GSEMO.

Homogeneous island model:

- There is an upper limit to the number of island still leading to linear speed-up
- Less dense topologies (i.e. grid, ring) assure linear speed-up for a smaller maximum number of islands but lead to lower communication effort.
- The migration probability gives a smooth trade-off between the maximum number of islands that yields to linear speedup and the communication effort

Heterogeneous island model:

- A simpler model providing the same linear speed-up with less comunication effort (with smart policy)
- Making the islands investigating different partitions of the search space might be a good idea