Influence of Replacement and Selection on Spurious Dependencies
Paper Reviews

- Influence of Selection and Replacement Strategies on Linkage Learning in BOA

- Spurious Dependencies and EDA Scalability
  - Elizabeth Radetic and Martin Pelikan (2010)
Influence of Selection and Replacement Strategies on Linkage Learning in BOA

BOA and Spurious Dependencies

- The probabilistic models are learned from a sample of limited size, so particular features of the specific sample are also encoded.

- This is a well-known problem in machine learning, overfitting.
Among the dependency groups captured by the Bayesian network with decision trees, spurious linkages are included.

The conditional probabilities nearly express independency between the spurious variables and the correct linkage.

This kind of overfitting does not greatly affect the capability of sampling such variables.
However, several efficiency enhancement techniques for EDAs crucially rely on the structural accuracy of the probabilistic models.

One example is the exploration of substructural (BB-wise) neighborhoods for local search in BOA (Lima et al., 2006).
- While significant speedups were obtained by incorporating model-based local search, the scalability of this speedup decreased for larger problem sizes due to overly complex model structures.

This paper investigates the influence of different selection and replacement strategies on the quality of BOA models.
Test problem: \( m-k \) trap function

For accurate linkage learning in BOA at least one of the variables of each trap subfunction should depend on all remaining \((k - 1)\) variables.
- \((X_1 \leftarrow X_2, X_3, X_4, X_5)\) encodes a linkage group between all variables for the first 5-bit trap subfunction.
- For \((X_1 \leftarrow X_2, X_3, X_4, X_5, X_6, X_{11})\), \(X_6\) and \(X_{11}\) are spuriously linked variables.

At each generation four different measures are analyzed taking into account only dependency groups of order \( k \) or higher:
- Proportion of BBs with correct linkage group
- Proportion of BBs with spurious linkage group
- Proportion of BBs with a linkage group
- Average size of spurious linkage
The qualitative results agree with a study (Yu et al. 2007) about the influence of selection pressure in the population size requirements for entropy-based EDAs.

Figure 1: (a) Population size and (b) number of function evaluations required for different tournament sizes to solve concatenated 5-bit traps of varying string length \( \ell \).
The speedup decreases with increasing problem size, suggesting that this is not a scalable speedup.

Figure 2: Speedup obtained when using the optimal tournament size for each problem size compared to the standard binary tournament \((s = 2)\).
Influence of the Selection Method

- Equivalent tournament size ($s$) and truncation threshold ($\tau$) for the same selection intensity ($I$).

<table>
<thead>
<tr>
<th>$I$</th>
<th>$s$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.56</td>
<td>2</td>
<td>66%</td>
</tr>
<tr>
<td>0.84</td>
<td>3</td>
<td>47%</td>
</tr>
<tr>
<td>1.03</td>
<td>4</td>
<td>36%</td>
</tr>
</tbody>
</table>
Influence of the Selection Method

Figure 3: Linkage group information captured by the probabilistic model of BOA along the run for different tournament sizes, \( s = \{2, 3, 4\} \), when solving \( m = 24 \) concatenated traps of order \( k = 5 \) (\( \ell = 120 \)). Tournament selection and full replacement is used.
Influence of the Selection Method

Figure 4: Linkage group information captured by the probabilistic model of BOA along the run for different truncation thresholds, \( \tau = \{66\%, 47\%, 36\%\} \), when solving \( m = 24 \) concatenated traps of order \( k = 5 \) (\( \ell = 120 \)). Truncation selection and full replacement is used.
Influence of the Selection Method

- For tournament selection:
  - Selection is performed at the individual-level rather than at the BB level. Therefore, a top individual does not necessarily have mostly good substructures.
  - This induces the learning process of BNs into some uncertainty, that is more pronounced in the initial generations.

- For truncation selection:
  - The results are significantly better in terms of structural accuracy.
If we compare tournament selection with $s = 2$ and truncation selection with $\tau = 36\%$, the requirements for truncation are smaller.

Figure 5: (a) Population size and (b) number of function evaluations required for different selection strategies when solving concatenated 5-bit traps of varying total string length $\ell$. Full replacement is used. Although truncation selection requires larger population sizes, using tournament selection with equivalent population sizes to those required by truncation does not significantly improve the linkage information.
Between truncation selection and tournament selection:

- Truncation selection has a higher loss of diversity (for the same selection intensity) which would lead to premature convergence.
- A faster and clear distinction between good individuals and just above average individuals reduces the noise faced by the learning process of BNs.
- While in tournament selection the number of copies of an individual is proportional to its rank, in truncation no particular relevance is given to very good individuals.

Figure 6: Average size of the spurious linkage for different selection strategies when solving concatenated 5-bit traps of varying total string length $\ell$. Full replacement is used.
Figure 7: Linkage group information captured by the probabilistic model of BOA along the run for different replacement methods when solving $m = 24$ concatenated traps of order $k = 5$ ($\ell = 120$). ER 50% stands for the replacement of the worst 50% parents, FR for full replacement, and RTR for restricted tournament replacement. Binary tournament selection is used.
Influence of the Replacement Method

Figure 8: (a) Population size and (b) number of function evaluations required for different replacement strategies when solving concatenated 5-bit traps with varying total string length $\ell$. Binary tournament selection is used.
Influence of the Replacement Method

- **RTR:**
  - By preserving diversity in the population, it requires smaller population size and fewer evaluations.
  - The quality of the model is not the best, because the drawback of using tournament selection is aggravated with smaller population sizes and increased diversity due to niching.

![Graph showing average size of spurious linkage for different replacement strategies](image)

**Figure 9:** Average size of the spurious linkage for different replacement strategies when solving concatenated 5-bit traps with varying total string length $\ell$. Binary tournament selection is used.
For the purpose of having accurate structural linkage information:
- Using truncation instead of tournament selection is much better.
- The replacement strategy was found to be relevant only if tournament selection is used, in which case the full replacement of the parents by their offspring is the most appropriate strategy.

If overall performance (NFE) is our main concern:
- Tournament selection and RTR are clearly the best options.

This paper provides important information about the trade-off between the consequent computational cost and the degree of accuracy for the learned linkage information in BOA.
Spurious Dependencies and EDA Scalability

Elizabeth Radetic and Martin Pelikan (2010)
Yu et al. (2005) investigated the effects of spurious dependencies on convergence time in EDAs.

Lima et al. (2007) noted that the inaccuracy in EDA models can negatively impact the performance of efficiency enhancement techniques such as substructural hillclimbing.

Nevertheless, little work exists that studies the effects of spurious dependencies on EDA scalability.
Sastry and Goldberg (2000) explained that it is due to the bias in the original population.

The initial bias is due to the finite population size.

Additional spurious dependencies are discovered (Pelikan et al., 2002) when the population size becomes too large. This is because selection introduces statistical dependencies.
This paper uses MPM models of ECGA to study the effects of spurious dependencies.

The order of linkage groups is tuned by specifying their size using the parameter $k_{\text{spurious}}$. (Assume that all the linkage group are of the same size.)

Ideally, $k_{\text{spurious}}$ should be 1 (the model is simply the probability vector) for the onemax model.
EDA Scalability: Population Sizing

- The initial-supply population-sizing model:
  \[ N = \chi^k (k \ln \chi + \ln m) \]
  where \( \chi \) is the cardinality of the alphabet.

- With respect to the onemax model for spurious dependencies:
  \[ N = 2^{k_{spurious}} \left( k_{spurious} \ln 2 + \ln \frac{\ell}{k_{spurious}} \right) \]

- The ratio of the population size required for an arbitrary value of \( k_{spurious} \geq 1 \) and that for the probability vector:
  \[ \gamma_{is} = 2^{k_{spurious} - 1} \frac{k_{spurious} \ln 2 + \ln \frac{\ell}{k_{spurious}}}{\ln 2 + \ln \ell} \]
EDA Scalability: Population Sizing

- The gambler’s ruin model: \[ N = -2^{k-1} \ln(\alpha) \frac{\sigma_{BB} \sqrt{\pi m'}}{d} \]
  - where \( \alpha \) is the error, \( \sigma_{BB} \) is the standard deviation of the fitness of one subproblem, \( m' = m - 1 \), and \( d \) is the signal.

- For onemax and EDAs based on probability vectors, \( \sigma_{BB} = 0.5 \) and \( d = 1 \):
  \[ N = -\frac{1}{2} \ln(\alpha) \sqrt{\pi (\ell - 1)} \]

- Assuming an MPM model with the linkage group size \( k_{spurious} \):
  \[ N = -2^{k_{spurious}-2} \ln(\alpha) \sqrt{\pi (\ell - 1)} \]

- The ratio by which the decision-making population size bound increases with \( k_{spurious} \) can be estimated as:
  \[ \gamma dm = 2^{k_{spurious}} \]
Yu et al. (2007) further refined the model for population sizing in entropy-based model building EDAs as \( \Omega(2^{2k} \ell \ln \ell) \).

The estimates which take model building into consideration grow faster than both the \( \Omega(2^k \ln \ell) \) bound for the initial supply and the \( \Omega(2^k \ell^{1/2}) \) bound for the decision making.

Nonetheless, for the purposes of this paper, the population sizing for the model building is not relevant.
EDA Scalability: Convergence Time

- For EDAs with the probability vector on onemax, the convergence time can be estimated by:
  \[ T = \left( \frac{\pi}{2} - \arcsin(2p - 1) \right) \frac{\sqrt{\ell}}{T} \]
  - where \( p \) is the initial proportion of building blocks.

- It can be hypothesized that the time to convergence for the onemax model for spurious dependencies will still be upper bounded by this equation.
  - A model with spurious dependencies does not enforce statistical independence of pairs of string positions. Because of this, the populations may lose diversity faster than with the probability vector.

- The only scenario that contradicts this hypothesis is that with niching (RTR).
EDA Scalability: Discussion

- This paper provides a conservative bound on the impact of spurious linkage on population sizing and time to convergence.
  - This paper uses the onemax model.
  - In practice, many of the spurious dependencies can be expected to be eliminated over time.

- The population size is still expected to be $\Omega(2^{2k} \ell \ln \ell)$, which outweighs the effects of spurious linkage.
Experiments

- Two sets of experiments were conducted:
  - Standard ECGA was used to solve onemax, and its models were analyzed to determine the types and quantities of spurious dependencies discovered.
  - An EDA with a fixed model structure was used to solve onemax, and the effects of spurious linkage on performance with the special focus on the population size were measured.
  - Each run was terminated when one solution of the desired quality had been found whereas the theory assumed full convergence.
Experiments

(a) Number of spurious linkage groups  (b) Avg. size of spurious linkage groups  (c) Average linkage group size

Figure 1: The average number of spurious linkage groups (groups of size $\geq 2$), the average size of linkage groups of size $\geq 2$, and the average linkage group size (including all linkage groups) for ECGA on onemax. Three replacement strategies are considered: full replacement, elitist replacement and RTR. For each problem size and replacement strategy, the results represent an average over 100 runs (10 bisections of 10 runs each).
Figure 2: Growth of the population size with respect to the group size for a problem of 300 bits. The left-hand side shows the actual population sizes compared to the theoretical model, whereas the right-hand side shows the ratio of the population sizes with spurious linkage and the population sizes with no spurious linkage.
Experiments

Figure 3: Growth of the population size with respect to the spurious linkage group size.
Yu et al. (2005) suggested that spurious linkage undermines the ability of EDAs to effectively mix partial solutions.

Figure 4: Growth of the number of generations with respect to the spurious linkage group size.
Experiments

Figure 5: The population size, the number of generations, and the number of evaluations for the 300-bit onemax with varying numbers of spurious linkage groups of size 2 (the remaining groups are of size 1).
Spurious dependencies were shown to increase the population size required to obtain a solution of sufficient quality.

Only in the presence of niching, the negative effects of spurious dependencies were empirically shown to be substantial.

Many research challenges remained for future work:
- Combining the results on the effects of spurious dependencies with those on model building.
- Examining the effects of spurious dependencies for models in which the subproblems overlap (BN can be used as the base class of models).

The study of the interaction between models with spurious dependencies and niching is a key topic for future work in EDAs.
My Opinions

- These two papers have good background introductions and conclusions.

- Any insight into improving the selection and replacement (niching) methods?