Estimation of Distribution Algorithms based on Copula Functions

Rogelio Salinas-Gutiérrez
Department of Computer Science
Center for Research in Mathematics (CIMAT)
Guanajuato, México
rsalinas@cimat.mx

Arturo Hernández-Aguirre†
Department of Computer Science
Center for Research in Mathematics (CIMAT)
Guanajuato, México
artha@cimat.mx

Enrique R. Villa-Diharce‡
Department of Probability and Statistics
Center for Research in Mathematics (CIMAT)
Guanajuato, México
villadi@cimat.mx

ABSTRACT
The main objective of this doctoral research is to study Estimation of Distribution Algorithms (EDAs) based on copula functions. This new class of EDAs has shown that it is possible to incorporate successfully copula functions in EDAs.

Categories and Subject Descriptors
I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search—Heuristic methods; G.1.6 [Numerical Analysis]: Optimization—Global optimization, Unconstrained optimization

General Terms
Algorithms, Design, Performance

Keywords
EDAs, copula functions, graphical models

1. INTRODUCTION
Estimation of Distribution Algorithms (EDAs) are a well established paradigm in Evolutionary Computation (EC). This class of evolutionary algorithms employs probabilistic models for searching and generating promising solutions. The goal in EDAs is to take into account the dependence structure in the best individuals and transfer them into the next population. A pseudocode for EDAs is shown in Algorithm 1.

Algorithm 1 Pseudocode for EDAs
1: assign $t \leftarrow 0$
2: generate the initial population $P_0$ with $N$ individuals at random
3: select a collection of $M$ solutions $S_t$, with $M < N$, from $P_t$
4: estimate a probabilistic model $M_t$ from $S_t$
5: generate the new population by sampling from the distribution of $S_t$
6: assign $t \leftarrow t + 1$
7: if stopping criterion is not reached go to step 2

The research on EDAs has been conducted in proposing and enhancing probabilistic models. Most of the probabilistic models used in EDAs for discrete domains are based on graphical models such as Bayesian and Markov Networks [9, 2, 24, 17, 16, 21, 22]. The discrete EDAs have been widely studied and they have been also used in several applications. However, for continuous domains, the Gaussian distribution is the most used assumption over the probabilistic model [12, 13, 14, 6]. For this reason, one of the current challenges for designing new continuous EDAs is finding multivariate models to adequately represent dependencies among the decision variables.

On the other hand, during the last decade, the copula functions have been widely used in many applications where nonlinear dependencies arise among variables. By using copula functions it is possible to separate the effect of dependence from the effect of marginal distributions in a joint distribution. However, they have been barely used in computer science applications where nonlinear dependencies are common and need to be represented.

The following section presents a brief introduction to copula functions and some EDAs based on copula theory.

2. CURRENT RESEARCH
One motivation for this research has been the possibility of proposing new EDAs for continuous domains. The copula based EDAs are able, at the same time, (1) to model dependencies among variables, and (2) to use flexible marginal distributions. The copula theory gives the means for getting such EDAs.
2.1 Copula Theory

Definition 1. A copula is a joint distribution function of standard uniform random variables. That is,

$$C(u_1, \ldots, u_d) = \Pr[U_1 \leq u_1, \ldots, U_d \leq u_d],$$

where $U_i \sim U(0,1)$ for $i = 1, \ldots, d$.

The following result, known as Sklar’s theorem, gives the relevance and practical utility to copula functions.

Theorem 1 (Sklar). Let $F$ be a $d$-dimensional distribution function with marginals $F_1, F_2, \ldots, F_d$, then there exists a copula $C$ such that for all $x$ in $\mathbb{R}^d$,

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

where $\mathbb{R}$ denotes the extended real line $[-\infty, \infty]$. If $F_1(x_1)$, $F_2(x_2)$, $\ldots$, $F_d(x_d)$ are all continuous, then $C$ is unique. Otherwise, $C$ is uniquely determined on $\text{Ran}(F_1) \times \text{Ran}(F_2) \times \cdots \times \text{Ran}(F_d)$, where Ran stands for the range.

According to Theorem 1 and using the chain rule for differentiating composite functions, any $d$-dimensional density $f$ can be represented as

$$f(x_1, \ldots, x_d) = c[F_1(x_1), \ldots, F_d(x_d)] \cdot \prod_{i=1}^{d} f_i(x_i),$$

where $c$ is the density of the copula $C$, and $f_i(x_i)$ is the marginal density of variable $x_i$.

Equation (1) shows that the dependence structure is modeled by the copula function. This expression separates any joint density function into the product of the multivariate copula density and the marginal densities. This contrasts with the usual way to model multivariate distributions, which suffers from the restriction that the marginal distributions are usually of the same type. The separation between marginal distributions and a dependence structure explains why the copula functions are suitable tools for modeling dependencies in a joint distribution.

The copula theory was introduced by Sklar [23] and has been developed during the last fifty years. The interested reader of this theory is referred to [11, 15, 25].

2.2 Related works

To the best of our knowledge the theses [3, 1] are the first attempts to incorporate a multivariate Gaussian copula function in EDAs. Since then, other related works have been published. These papers present EDAs based on (1) the Gaussian copula function [30, 28], and (2) Archimedean copula functions with a fixed dependence parameter [29, 26, 27, 7, 31]. Unlike the previous papers that use Archimedean copula functions, the work [10] presents a way of estimating the copula parameter. These works use copula functions to model pairwise relationships among all variables and do not employ graphical models. On the other hand, we have proposed the use of the maximum likelihood method and copula entropies in order to (1) estimate the copula parameters, and (2) build a graphical model which establishes the most important dependencies between variables [19, 20, 18].

2.3 Our proposed EDAs

Our research has been conducted for designing new continuous EDAs. These EDAs are described below.

- In [19] we have presented an EDA based on a chain graphical model and bivariate copula functions. Every chain link in the graphical model represents the dependence between two decision variables. A bivariate copula function is used for modeling such dependence. The relationship between the mutual information and the bivariate copula entropy [8] is used for measuring the dependence between variables. The probabilistic model used in this paper can be seen as a generalization of the well known EDA MIMIC$^2$ [12, 13]. In this work, for the first time, the estimation of copula parameters by means of the maximum likelihood function is presented.

- In [20], a regular vine [4, 5] is considered as a graphical model for designing a new EDA. This EDA is based on a particular family of regular vines called D-vine. In this work we propose a greedy algorithm for building such graphical model and we also show the theoretical relationship between the entropy of a trivariate copula function and the conditional mutual information.

- The previous works consider the use of a fixed copula function. Based on this observation, the work [18] is a recent contribution for selecting copula functions. A tree graphical model is employed in this paper. A copula function is selected for modeling the dependence between variables in each tree branch. The copula function is selected from a set of six bivariate copula functions and it is selected according to the likelihood function.

Some results obtained by D-vine EDA for three well known test functions in ten dimensions are reported in Table 1. For the Rosenbrock problem, the range of values reached by a D-vine EDA is closer to the global minimum than the range of values of the other EDAs. This would mean that a dependence structure based on a D-vine EDA is more adequate than the chain structure of the MIMIC. More details about these results can be seen in [20].

Figure 1 shows recent results by using a copula selection procedure in EDAs. In this figure, the success rate is reported by two chain and two tree graphical models. These results show a clear advantage for probabilistic models that are incorporated with a copula selection procedure. More results are presented in [18].

A general procedure for estimating the graphical model in our work has been the minimization of the Kullback-Leibler divergence between the true density and the proposed density. This procedure has also used the natural relationship between the entropy of a bivariate copula function and the marginal mutual information of two variables [8].

3. CONCLUSIONS

The use of copula functions in EDAs has opened a new research area. Our approach has been based on copula functions whose parameters are estimated by means of maximum likelihood, and the most important dependencies are represented by a graphical model.
Table 1: Descriptive results for the fitness of three test functions.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Best</th>
<th>Median</th>
<th>Mean</th>
<th>Worst</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ackley</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-vine EDA</td>
<td>2.71E-005</td>
<td>4.11E-005</td>
<td>4.09E-005</td>
<td>5.16E-005</td>
<td>5.82E-006</td>
</tr>
<tr>
<td>MIMIC</td>
<td>2.89E-005</td>
<td>3.91E-005</td>
<td>3.92E-005</td>
<td>4.95E-005</td>
<td>5.21E-006</td>
</tr>
<tr>
<td>UMDA</td>
<td>2.83E-005</td>
<td>4.08E-005</td>
<td>4.15E-005</td>
<td>5.07E-005</td>
<td>4.63E-006</td>
</tr>
<tr>
<td><strong>Rosenbrock</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-vine EDA</td>
<td>0.20</td>
<td>5.75</td>
<td>5.28</td>
<td>9.21</td>
<td>2.70</td>
</tr>
<tr>
<td>MIMIC</td>
<td>0.91</td>
<td>6.40</td>
<td>6.42</td>
<td>13.48</td>
<td>2.72</td>
</tr>
<tr>
<td>UMDA</td>
<td>7.93</td>
<td>8.02</td>
<td>8.10</td>
<td>10.28</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Sphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-vine EDA</td>
<td>3.37E-007</td>
<td>7.75E-007</td>
<td>7.48E-007</td>
<td>9.99E-007</td>
<td>1.69E-007</td>
</tr>
<tr>
<td>MIMIC</td>
<td>4.48E-007</td>
<td>8.47E-007</td>
<td>8.14E-007</td>
<td>9.96E-007</td>
<td>1.92E-007</td>
</tr>
<tr>
<td>UMDA</td>
<td>4.20E-007</td>
<td>8.56E-007</td>
<td>8.11E-007</td>
<td>9.94E-007</td>
<td>1.64E-007</td>
</tr>
</tbody>
</table>

The obtained performance results suggest that the incorporation of copula functions is a real option for designing new continuous EDAs. Moreover, our reported experiments have shown that modeling adequately the dependencies between decision variables is a good mechanism for getting better performance results. Future work for our investigation will focus on the design of algorithms with strategies for enhancing their performance.

4. ACKNOWLEDGMENTS

The first author acknowledges support from the National Council of Science and Technology of México (CONACyT) through a scholarship to pursue graduate studies in the Department of Computer Science at the Center for Research in Mathematics.

5. REFERENCES


![Figure 1: Success rate results for (a) a separable function, and (b) a non-separable function. The horizontal axis represents the dimension problem and the vertical axis represents the success rate. The solid line is used for the EDAs based on copula selection and the dashed line for the EDAs based on the Gaussian copula.](image)