A Mathematical Analysis of EDAs with Distance-based Exponential Models

Imanol Unanue
University of the Basque Country
UPV/EHU
Donostia-San Sebastian, Spain
imanol.unanue@ehu.eus

María Merino
University of the Basque Country
UPV/EHU
Leioa, Spain
maria.merino@ehu.eus

Jose A. Lozano
Basque Center for Applied
Mathematics (BCAM)
University of the Basque Country
UPV/EHU
ja.lozano@ehu.eus

ABSTRACT

Estimation of Distribution Algorithms have been successfully used for solving many combinatorial optimization problems. One type of problems in which Estimation of Distribution Algorithms have presented strong competitive results are permutation-based combinatorial optimization problems. In this case, the algorithms use probabilistic models specifically designed for codifying probability distributions over permutation spaces. One class of these probability models is distance-based exponential models, and one example of this class is the Mallows model. In spite of the practical success, the theoretical analysis of Estimation of Distribution Algorithms for permutation-based combinatorial optimization problems has not been extensively developed. With this motivation, this paper presents a first mathematical analysis of the convergence behavior of Estimation of Distribution Algorithms based on the Mallows model by using an infinite population to associate a dynamical system to the algorithm. Several scenarios, with different fitness functions and initial probability distributions of increasing complexity, are analyzed obtaining unexpected results in some cases.

CCS CONCEPTS

• **Mathematics of computing → Mathematical analysis**; **Mathematical optimization**;

KEYWORDS

Estimation of Distribution Algorithm, Permutation-based Combinatorial Optimization Problems, Mathematical Modeling, Theoretical Analysis, Mallows model

1 INTRODUCTION

Estimation of Distribution Algorithms (EDAs) [6, 9], is a class of Evolutionary Algorithms introduced by Mühlenbein and Paaβ [8]. The main characteristic of EDAs is that they learn a probability distribution from a database containing the selected solutions from the previous generation at each iteration. The new set of solutions is sampled from the learned probability distribution.

Recently, EDAs have been successfully used to solve permutation-based combinatorial optimization problems [1, 10]. In order to do that, these EDAs use probabilistic models specifically designed for codifying probability distributions over permutation spaces. In particular, the authors of [1] used a Mallows model. This model can be included in a more general class of probability models: distance-based exponential models. However, it is still not clear which mechanisms allow these algorithms to obtain these results.

Similar to Genetic Algorithms, the first theoretical studies on EDAs were focused on the convergence behavior of algorithms such as UMDA [11] and PBIL [4]. Nonetheless, several works have been presented recently in the literature with the aim of attaining new theoretical results about the runtime, the population sizing or the model accuracy of EDAs. For a current state-of-the-art, see [5]. However, the previous theoretical studies are designed for binary and continuous search spaces.

Our objective in this work is to present a mathematical model to analyze the behavior of an EDA based on a Mallows model in some scenarios with different fitness functions and initial probability distributions of increasing complexity, and obtain the first theoretical results over the permutation space.

2 THE MALLOWS MODEL

The Mallows model [7] is a distanced-based exponential probability model over permutations, considered as the analogous of the Gaussian distribution. The probability value of every permutation $\sigma \in \Sigma_n$ depends on two parameters: a central permutation σ_0 and a spread parameter θ. The Mallows model is defined as

$$P(\sigma) = \frac{1}{\varphi(\theta)} e^{-\theta d(\sigma, \sigma_0)} \tag{1}$$

where $d(\sigma, \sigma_0)$ is the distance from σ to σ_0, and $\varphi(\theta)$ is the normalization constant. By definition, any two solutions at the same distance from the central permutation have the same probability.

The most used distance in the literature for Mallows model is Kendall's-τ distance, and we will use it during this work. Kendall's-τ distance $d(\sigma, \pi)$ is the minimum number of adjacent transpositions needed to bring σ^{-1} to π^{-1}.

3 EDAS BASED ON MALLOWS MODELS

An EDA can be considered as a sequence of probability distributions, each one given by a stochastic transition rule $\mathcal{G}$. So, $P_i = \mathcal{G}(P_{i-1}) = \mathcal{G}^i(P_0)$, $\forall i \in \mathbb{N}$. Hence, the convergence behavior is described as follows: $\lim_{i \longrightarrow \infty} \mathcal{G}^i(P_0)$.

The application of the EDA schema to deal with optimization problems can involve an unapproachable variety of situations and behaviors. To study the behavior of an EDA, we assume that the population is infinite [2]. In EDAs with infinite populations, the empirical probability distribution induced by the solutions in the current population and the selected population converge to their underlying probability distributions P_i and P_i^S, and they could be thought of as the population and the selected proportions of each individual at iteration i. At each iteration of the algorithm a probability distribution P_{i+1} is obtained (see Algorithm 2 from [2]).

Hence, $P_i = G(P_{i-1}) = G^i(P_0)$, where G returns the expected probability vector for the next iteration: $G(P_i) = E[\mathcal{G}(P_i)]$. So, when the population size tends to infinity,

$$\lim_{i \longrightarrow \infty} \mathcal{G}^i(P_0) = \lim_{i \longrightarrow \infty} G^i(P_0). \tag{2}$$

In order to study the convergence behavior of the algorithm with infinite population, a composition of the selection operator ϕ and the approximation step a used to learn the probability distribution is used: $G = a \circ \phi$.

The selection operator used in this work has been the widely used 2-tournament selection (adapted to an infinite population):

$$p_i^S(\sigma) = 2 \sum_{\pi | f(\sigma) > f(\pi)} p_i(\sigma) p_i(\pi) + \sum_{\pi | f(\sigma) = f(\pi)} p_i(\sigma) p_i(\pi). \tag{3}$$

In addition, in our mathematical model, at each generation of the algorithm a new Mallows model will be learnt from P_i^S by using the maximum likelihood estimators of σ_0 and θ, $\hat{\sigma}_0$ and $\hat{\theta}$, adapted from [3] to an infinite population:

$$\hat{\sigma}_0 = \arg \min_{\sigma \in \Sigma_n} \sum_{\pi \in \Sigma_n} d(\pi, \sigma) \cdot p^S(\pi) \tag{4}$$

$$\sum_{\pi \in \Sigma_n} d(\pi, \hat{\sigma}_0) \cdot p^S(\pi) = \frac{n-1}{e^{\hat{\theta}} - 1} - \sum_{i=1}^{n-1} \frac{n-i+1}{e^{(n-i+1)\hat{\theta}} - 1}. \tag{5}$$

4 LIMITING BEHAVIOR IN SOME SCENARIOS

Our mathematical modeling is applied and proved to some scenarios. Each scenario is defined by a fitness function f and an initial probability distribution P_0. The different fitness functions f used for the scenarios have been the uniform function, *needle in a haystack* function and Mallows probability distribution. The following results have been proven:

- When f is a constant function, for any P_0 Mallows distribution, the algorithm keeps the same model forever.
- When f is a *needle in a haystack* function and P_0 a Mallows distribution centered in the optimal solution, the algorithm converges to the optimal solution.
- When f is a Mallows model centered in the optimal solution ($\theta > 0$) and P_0 a uniform distribution, the algorithm

converges to a Mallows model centered in the optimal distribution.

Moreover, when f is a Mallows model centered in the optimal solution ($\theta > 0$) and P_0 another Mallows model centered in a solution σ_0, we have made some conjectures:

- G operator can only estimate central permutations between the optimal solution and σ_0.
- The algorithm can only converge to solutions at distance $d \leq \lfloor n(n-1)/4 \rfloor$ (with a single exception).

5 CONCLUSIONS

We have presented a mathematical model to study an EDA with infinite populations using discrete dynamical systems and distance-based exponential models. Several problems have been presented and studied, combining the formal results and some conjectures (based on some experiments). In general, the presented theoretical model has shown that in most cases it converges to the optimal solution.

REFERENCES

[1] Josu Ceberio, Ekhine Irurozki, Alexander Mendiburu, and Jose A Lozano. 2014. A distance-based ranking model estimation of distribution algorithm for the flowshop scheduling problem. *IEEE Transactions on Evolutionary Computation* 18, 2 (2014), 286–300.

[2] Carlos Echegoyen, Roberto Santana, Alexander Mendiburu, and Jose A Lozano. 2015. Comprehensive characterization of the behaviors of estimation of distribution algorithms. *Theoretical Computer Science* 598 (2015), 64–86.

[3] Michael A Fligner and Joseph S Verducci. 1986. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)* (1986), 359–369.

[4] Cristina González, Jose A Lozano, and Pedro Larrañaga. 2000. Analyzing the population based incremental learning algorithm by means of discrete dynamical systems. *Complex Systems* 12 (2000), 465–479.

[5] Martin S Krejca and Carsten Witt. 2018. Theory of estimation-of-distribution algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 1170–1197.

[6] Pedro Larrañaga and Jose A Lozano. 2002. *Estimation of distribution algorithms: A new tool for evolutionary computation*. Vol. 2. Springer Science & Business Media.

[7] Colin L Mallows. 1957. Non-null ranking models. *Biometrika* 44, 1/2 (1957), 114–130.

[8] Heinz Mühlenbein and Gerhard Paaß. 1996. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*. Springer, 178–187.

[9] Martin Pelikan, Mark W Hauschild, and Fernando G Lobo. 2015. Estimation of distribution algorithms. In *Springer Handbook of Computational Intelligence*. Springer, 899–928.

[10] Shigeyoshi Tsutsui. 2006. Node histogram vs. edge histogram: A comparison of probabilistic model-building genetic algorithms in permutation domains. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*. IEEE, 1939–1946.

[11] Qingfu Zhang and Heinz Muhlenbein. 2004. On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on evolutionary computation* 8, 2 (2004), 127–136.