



Using metaheuristic algorithms for parameter estimation in generalized Mallows models



Juan A. Aledo^{a,*}, Jose A. Gámez^b, David Molina^a

^a Departamento de Matemáticas, Universidad de Castilla-La Mancha, Albacete 02071, Spain

^b Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete 02071, Spain

ARTICLE INFO

Article history:

Received 14 May 2014

Received in revised form

13 September 2015

Accepted 29 September 2015

Available online 21 October 2015

Keywords:

Generalized Mallows model

Parameter estimation

Metaheuristic algorithms

Search methods

ABSTRACT

This paper deals with the problem of parameter estimation in the generalized Mallows model (GMM) by using both local and global search metaheuristic (MH) algorithms. The task we undertake is to learn parameters for defining the GMM from a dataset of complete rankings/permutations. Several approaches can be found in the literature, some of which are based on greedy search and branch and bound search. The greedy approach has the disadvantage of usually becoming trapped in local optima, while the branch and bound approach, basically A^* search, usually comes down to approximate search because of memory requirements, losing in this way its guaranteed optimality. Here, we carry out a comparative study of several MH algorithms (iterated local search (ILS) methods, variable neighborhood search (VNS) methods, genetic algorithms (GAs) and estimation of distribution algorithms (EDAs)) and a tailored algorithm A^* to address parameter estimation in GMMs. We use 22 real datasets of different complexity, all but one of which were created by the authors by preprocessing real raw data. We provide a complete analysis of the experiments in terms of accuracy, number of iterations and CPU time requirements.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Although the problem of dealing with permutations is a long-studied topic, there has been remarkable interest in this field from the statistics and machine learning community in the last two decades. One reason for this is the increasing presence of problems whose basic data are rankings, which are represented as permutations (e.g. recommender systems, combinatorial optimization, preferences, etc.).

Of the different ways of dealing with permutations/rankings in a compact manner, probabilistic models have received more attention than others [14]. Undoubtedly, the most famous probabilistic model for permutations is the Mallows model (MM) [32], introduced by C.L. Mallows in the late fifties. This model has recently experienced a noteworthy revival, and has been applied to problems of different nature: combinatorial optimization [5], supervised classification [7], preference modeling and collaborative filtering [47], content modeling [6], etc.

The Mallows distribution is a distance-based ranking model which has a certain resemblance to the Gaussian distribution. It also belongs to the exponential family and is specified by two parameters: a permutation π_0 which can be seen as the consensus ranking (the mode), and a spread parameter θ which determines how concentrated the different rankings are around π_0 , that is, how flat or peaked the distribution is. Perhaps the most famous extension of the Mallows model is the generalized Mallows model (GMM) [13]. In the GMM, again we have the mode or consensus permutation, π_0 , but instead of a single spread parameter θ , GMM requires $n - 1$ spread parameters $\{\theta_j\}_{j=1}^{n-1}$, where each θ_j affects the j th position of the

permutation. When an appropriate distance is used, the GMM can be seen as a *multistage ranking model* [15].

In this paper, our aim is to estimate the parameters required in the GMM from a dataset of permutations/rankings by using a variety of widely tried and tested search-based metaheuristic (MH) methods. This task has been previously approached by using greedy heuristics and branch-and-bound (A^*) algorithms [2,33,35]. However, the results achieved for GMM parameters are far from being as good as the ones obtained when learning the parameters for the Mallows model. The main problem lies in the fact that in the Mallows model the estimation of the parameters can be performed independently, first the consensus permutation (which simply comes down to the well-known Kemeny ranking problem [27]), and then the spread value, while in the GMM constructive algorithms need to estimate the corresponding θ_j simultaneously with the consensus permutation that is being built step by step.

In this study we propose to approach the problem by using both *local* and *global* search metaheuristic algorithms. This type of method searches for the entire consensus permutation all at once, and not constructively item by item as the Borda and A^* algorithms do.

As mentioned above, the main difference between GMM and MM is that now the problem does not come down to the Kemeny ranking problem, because the quality of a candidate permutation does not depend only on itself but also on the estimated θ_j values. Therefore, we need to interleave the estimation of candidate permutations with the estimation of their corresponding set of θ_j values, which, given a permutation, can be efficiently carried out by using a numerical algorithm for convex optimization [33]. As the inference engine for guiding the search we propose to apply the following MHs methods:

- *Iterated local search (ILS) methods* [31], which are multi-start local search algorithms that try to escape from a current local optimum by perturbing it, and use

* Corresponding author. Tel.: +34 967599200; fax: +34 967599224.

E-mail addresses: juanangel.aledo@uclm.es (J.A. Aledo), jose.gamez@uclm.es (J.A. Gámez), david.molina@uclm.es (D. Molina).

the resulting configuration to seed a new hill climbing iteration. Regarding the hill climbing stage, we have considered two different neighborhoods.

- *Variable neighborhood search (VNS)* methods [21], which are also local search algorithms which alternate different neighborhoods to guide the search.
- *Genetic algorithms (GAs)*. When addressing optimization problems by using global search, evolutionary algorithms [37] and, in particular, GAs [20] have proven to perform successfully in permutation-based problems [1].
- *Estimation of distribution algorithms (EDAs)* [29]. These are a novel family of evolutionary computation algorithms based on modeling a population by using probabilistic models. Here we have considered two specific algorithms to deal with permutations: a node-based histogram EDA (NHM) and an edge-based histogram EDA (EHM).

Thus, we have carried out a comparative study of the selected MHs algorithms (ILS, VNS, GAs and EDAs) and the branch-and-bound (A*) algorithm, which was specifically designed to cope with the target problem by Ali and Meila [2,33,35]. We use 22 real datasets of different complexity, all but one of which were created by the authors by preprocessing real raw data (see Section 5.1).

As the experiments reported here demonstrate, the proposed MHs algorithms obtain better results (statistically significant) than the competing approaches considered. Obviously, MH algorithms need far more time than the Borda algorithm. However, given a dataset, the learning task is usually performed only once and so can be run off-line. On-line inference tasks will benefit from repeatedly using a good (fast) model. The use of MHs is widespread when the goal is to obtain accurate models at the expense of time consumption [10,17].

Our paper is organized as follows. In Section 2 we provide preliminaries and basic notations regarding rankings and the GMM. In Section 3 we review related/competing approaches to dealing with the problem of parameter estimation in GMM. Section 4 is devoted to describing the proposed MH algorithms. In Section 5 we describe our experimental study, detailing the datasets, methodology and results. We have designed two kinds of experiments by varying the number of evaluations that an algorithm is allowed to carry out, in order to test different features of the MH algorithms. Finally, in Section 6 we present our conclusions.

2. Preliminaries

In this section we introduce the notation as well as some background on the problem under study.

Suppose we have to rank n items labeled $1, 2, \dots, n$. Then, any permutation π of these items represents a ranking. The space of all the possible rankings forms the *symmetric group* \mathbb{S}_n :

Definition 1. (Symmetric group)[12] The symmetric group \mathbb{S}_n is the group whose elements are the permutations of the n symbols $\{1, 2, \dots, n\}$, and whose group operation is the composition of such permutations, which are treated as bijective functions from the set of symbols to itself. Since there exists $n!$ different permutations of size n , the order of the symmetric group is $n!$.

We will write $\pi = (x_1 x_2, \dots, x_n)$ to indicate that x_1 is ranked as the first item, x_2 as the second one, and so on. We will also denote by $\pi(j)$ the j th element of π .

To compare two given rankings or permutations, *distances* are used as the common tool. Although different distance measures are available in the literature, the *Kendall tau* distance is usually considered for the definition of Mallows and generalized Mallows distributions (see Sections 2.1 and 2.2).

Definition 2. (Kendall distance [28]) The Kendall distance $d(\pi, \sigma)$ between two rankings $\pi, \sigma \in \mathbb{S}_n$ is defined as the total number of item pairs over which they disagree. There is disagreement over an item pair (i, j) , $1 \leq i < j \leq n$, if the relative order of i and j is different in π and σ . Hence

$$d(\pi, \sigma) = \sum_{i < j} 1_{[j <_{\sigma} i]}, \tag{1}$$

where $1_{[\cdot]}$ is the indicator function and $i <_{\pi} j$ means that i precedes j in the permutation π .

More precisely,

$$d(\pi, \sigma) = |\{(i, j) : i < j, (\sigma(i) < \sigma(j) \wedge \pi(i) > \pi(j)) \vee (\sigma(i) > \sigma(j) \wedge \pi(i) < \pi(j))\}|$$

Given two permutations $\pi, \sigma \in \mathbb{S}_n$, $V_j(\pi, \sigma)$, $j = 1, \dots, n - 1$, is defined as [33]:

$$V_j(\pi, \sigma) = \sum_{i=j+1}^n 1_{[\pi(i) <_{\sigma} \pi(j)]} \tag{2}$$

Thus, V_1 stands for the number of elements ranked in π after $\pi(1)$ and ranked in σ before $\pi(1)$; V_2 corresponds to the number of elements ranked in π after $\pi(2)$ and ranked in σ before $\pi(2)$, and so forth (see [13] for more details).

Note that V_j takes values in $\{0, \dots, n - j\}$. Moreover, it is possible to determine π from σ and the $n - 1$ integer values of $V_1(\pi, \sigma), V_2(\pi, \sigma), \dots, V_{n-1}(\pi, \sigma)$ [35].

Using (1) and (2), the Kendall distance $d(\pi, \sigma)$ can be expressed in terms of the functions $V_j(\pi, \sigma)$ as

$$d(\pi, \sigma) = \sum_{j=1}^{n-1} V_j(\pi, \sigma).$$

When a set of permutations $\Pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ is considered, a *precedence matrix* $Q = [Q_{ij}]_{i,j=1:n}$ is useful in order to compute the *consensus* permutation, that is, the Kemeny ranking of Π [27]. Specifically, Q is computed as

$$Q_{ij} = \frac{1}{N} \sum_{k=1}^N 1_{[i <_{\pi_k} j]}. \tag{3}$$

In other words, Q_{ij} represents the fraction of times that item i is ranked before item j across all the N rankings.

Moreover, given $\sigma \in \mathbb{S}_n$ we define $\bar{V}_j(\Pi, \sigma)$ as the average of the $V_j(\pi_i, \sigma)$'s, that is,

$$\bar{V}_j(\Pi, \sigma) = \frac{\sum_{i=1}^N V_j(\pi_i, \sigma)}{N}.$$

Observe also (see [33]) that

$$\bar{V}_j(\Pi, \sigma) = \sum_{i=j+1}^n Q_{\sigma(i)\sigma(j)}. \tag{4}$$

2.1. The Mallows model

The Mallows model (MM) [32] is a distance-based probability distribution over permutation spaces which belongs to the exponential family. Given a distance over permutations, it can be defined by two parameters: the central permutation π_0 , and the spread parameter θ .

Definition 3. (Mallows model [32]) The Mallows model is the probability distribution that satisfies, for all rankings $\pi \in \mathbb{S}_n$,

$$P(\pi) = \frac{e^{-\theta \cdot d(\pi, \pi_0)}}{\psi(\theta)},$$

where the ranking π_0 and $\theta \geq 0$ are the model parameters, and $\psi(\theta)$ is a normalization constant (see [13]).

The parameter θ of the MM quantifies the concentration of the distribution around its peak π_0 . For $\theta > 0$, the probability of π_0 is the one with the highest value and the probability of the other $n! - 1$ permutations decreases with the distance from π_0 and the spread parameter θ . For $\theta = 0$ the uniform distribution is obtained. Due to this behaviour, the Mallows distribution on the space of permutations is considered analogous to the Gaussian distribution on the space of real numbers.

From now on, we assume that the Kendall distance is considered.

Download English Version:

<https://daneshyari.com/en/article/494866>

Download Persian Version:

<https://daneshyari.com/article/494866>

[Daneshyari.com](https://daneshyari.com)