Unsupervised Ranking of Multi-Attribute Objects Based on Principal Curves

Chun-Guo Li, Member, IEEE, Xing Mei, and Bao-Gang Hu, Senior Member, IEEE

Abstract—Unsupervised ranking faces one critical challenge in evaluation applications, that is, no ground truth is available. When PageRank and its variants show a good solution in related objects, they are applicable only for ranking from link-structure data. In this work, we focus on unsupervised ranking from multi-attribute data which is also common in evaluation tasks. To overcome the challenge, we propose five essential meta-rules for the design and assessment of unsupervised ranking approaches: scale and translation invariance, strict monotonicity, compatibility of linearity and nonlinearity, smoothness, and explicitness of parameter size. These meta-rules are regarded as high level knowledge for unsupervised ranking tasks. Inspired by the works in [12] and [35], we propose a ranking principal curve (RPC) model, which learns a one-dimensional manifold function to perform unsupervised ranking tasks on multi-attribute observations. Furthermore, the RPC is modeled to be a cubic Bézier curve with control points restricted in the interior of a hypercube, complying with all the five meta-rules to infer a reasonable ranking list. With control points as model parameters, one is able to understand the learned manifold and to interpret and visualize the ranking results. Numerical experiments of the presented RPC model are conducted on two open datasets of different ranking applications. In comparison with the state-of-the-art approaches, the new model is able to show more reasonable ranking lists.

Index Terms—Unsupervised ranking, multi-attribute, meta-rules, data skeleton, principal curves, Bézier curves

1 INTRODUCTION

Ranking is a fundamental problem in many applications, such as information retrieval and social choice. From the viewpoint of machine learning, ranking can be performed in an either supervised or unsupervised way as shown in Table 1. When supervised ranking [3] is able to evaluate the ranking performance from the given ground truth, unsupervised ranking seems more challenging because no ground truth label is available. Modelers or users will encounter a more difficult issue below:

“How can we insure that the ranking list from the unsupervised ranking is reasonable or proper?”

Although many unsupervised ranking strategies appeared recently, they took supervised evaluation measures like normalized discounted cumulative gain (NDCG) [4], mean average precision (MAP) [18], and recall precision of top-k lists. For example, unsupervised rank aggregation methods for meta-search are evaluated on LETOR [13] with NDCG and MAP. However, in cases of other ranking types that there is no ground-truth label, like countries and journals ranking, unsupervised ranking are still stuck in problems of evaluation.

PageRank [9] is one of accepted unsupervised ranking methods for web search since it makes full use of domain knowledge of web links, following the principle of “let the data speak for themselves” [33]. However, unsupervised ranking from multi-attribute objects does not have the same domain knowledge of web links such that PageRank is not suitable any more. Yet it highlights a possible solution. Unsupervised ranking from multi-attribute objects also has domain knowledge which can be made use of. In this work, we formally formulate this domain knowledge, which has been widely used in ranking problems (both supervised and unsupervised), into five essential meta-rules (or preferred features). These meta-rules can act not only as a high level evaluation of unsupervised ranking lists, but also as a guidance of designing ranking functions (Fig. 1).

Unsupervised ranking strategies for multi-attribute objects can be broadly divided into two categories: rank aggregation and structure based. Borda’s methods [19] and multinomial preference model (MPM) [21] fall into the first category which aggregate several ranking lists. Although Borda’s methods are simple to carry out, they are regardless of the specific distributions of different datasets. MPM pays attention to the score distribution, but it is an implicit rank aggregation method such that it hardly tells whether it is order-preserving (see Definition 2) which is an important requirement for ranking [11], [34]. The first principal component analysis (PCA) [17] and elastic map (Elmap) [28], fall into another category which try to depict the data skeleton as “ranking coordinate”. However, PCA is suitable for linearly-shaped datasets but not nonlinearly-shaped ones (e.g., Fig. 5 in Section 5) while Elmap has no guarantee of order-preserving. Yet, replacing Elmap directly by the other

---

X. Mei and B.-G. Hu are with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 95 ZhongGuanCun East Road, Beijing 100190, P.R. China. E-mail: {xmei, hubg}@nlpr.ia.ac.cn.

C.-G. Li is with the National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, 95 ZhongGuanCun East Road, Beijing 100190, P.R. China, and the College of Mathematics and Information Science, Hebei University, 180 Wusi East Road, Baoding, Hebei 071002, P.R. China. E-mail: licg@hbu.edu.cn.

Manuscript received 25 Jan. 2014; revised 26 Apr. 2015; accepted 14 May 2015. Date of publication 3 June 2015; date of current version 3 Nov. 2015. Recommended for acceptance by G. Das. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TKDE.2015.2441692
principal curve models cannot meet meta-rules either since they are not modeled specially for ranking tasks which requires order-preserving strongly. Example 1 shows the necessity of order-preserving which was neglected by most methods (e.g., [12], [19], [21]).

Example 1. Suppose we want to evaluate life qualities of countries with a principal curve based on two attributes: LEB\(^1\) and GDP.\(^2\) Each country is a data point in the two-dimensional plane of LEB and GDP. If the principal curve is approximated by a polyline as in Fig. 2a, the piece of the horizontal line is not strictly monotone. It makes the same ranking solution for \(x_1 = (58, 1.4)\) and \(x_2 = (58, 16.2)\) but \(x_2\) should be ranked higher than \(x_1\). For a general principal curve like the curve in Fig. 2b which is not monotone, two pairs of points are ordered unreasonably. The pair, \(x_3 = (74, 40.2)\) and \(x_4 = (82, 40.2)\), are put in the same place of the ranking list since they are projected to the same point which has the vertical tangent line to the curve. But \(x_4\) should be ranked higher for its higher LEB than \(x_3\). Another pair, \(x_5 = (75, 62.5)\) and \(x_6 = (81, 64.8)\), are also put in the same place but apparently \(x_6\) should be ranked higher than \(x_5\).

The following points highlight the main contributions of this paper:

- We propose five meta-rules for unsupervised ranking, which serve as high-level guidance in the design and assessment of unsupervised ranking approaches for multi-attribute objects. We justify that the five meta-rules are essential in model design and model selection for unsupervised ranking, but unfortunately some of them were overlooked by most of unsupervised ranking approaches.

- A ranking principal curve (RPC) model is presented for unsupervised ranking from multi-attribute numerical observations of objects. The presented model, parameterized by a cubic Bézier curve, can satisfy all of five meta-rules for ranking tasks, while most existing approaches fail to follow them. What’s more, the parameterization makes it more easier to interpret and visualize of ranking results.

- We develop the RPC learning algorithm, and theoretically prove the existence of a RPC and convergence of learning algorithm for given multi-attribute objects for ranking. With RPC learning algorithm, reasonable ranking lists for openly accessible data illustrate the good performance of the proposed unsupervised ranking approaches.

1.1 Related Works

For the evaluation of supervised ranking performance, NDCG \([4]\) and MAP \([18]\) are two widely used indicators in web search which involve the target ranking labels. But for the evaluation of unsupervised ranking performance where no target labels are available, there is not yet an acceptable evaluation method. Among existing unsupervised ranking methods \([19], [21], [23]\), most of them focus on the search ranking and are evaluated on two datasets of query searching results, TREC and LETOR, with NDCG and MAP. It is domain specific. For other types of unsupervised ranking problems, like countries and journals ranking, how to produce an acceptable ranking result remains a challenge.

The most famous unsupervised ranking is PageRank algorithm which makes full use of backlinks to rate a website for searching \([9]\). After that, numerous ranking algorithms are developed for web search including heuristic rank aggregation methods \([22]\). Most of these methods are unsupervised. Ranking on manifolds has provided a new ranking framework \([12], [14], [15], [16]\), which is different from general ranking functions such as ranking aggregation \([23]\). As one-dimensional manifolds, principal curves are able to perform unsupervised ranking tasks from multi-attribute numerical observations of objects \([12]\). But not all principal curve models can serve as ranking functions. For example, Elmap can well portray the contour of a molecular surface \([28]\) but would bring about a biased ranking list due to no guarantee of order-preserving \([12]\).

Domain knowledge has been widely used in machine learning fields. For example, ranking knowledge of monotonicity has been used to improve ordinal classification and regression \([10], [11], [34]\). For unsupervised learning, learning models with constraints of domain knowledge are expected to produce closer learning results to ground truths.
than those without knowledge constraints. An example is unsupervised image segmentation which has used the domain knowledge in segment assessing [24]. For unsupervised ranking tasks, their domain knowledge (like monotonicity) can also help to evaluate and design ranking functions which is the motivation of this work.

1.2 Paper Organization
The rest of this paper is organized as follows. Preliminaries and backgrounds are formalized in Section 2 and Section 3 respectively. In Section 4, five meta-rules are elaborated for unsupervised ranking. In Section 5, the RPC model is defined and formulated with a cubic Bézier curve following all the five meta-rules. RPC learning algorithm is designed in Section 6. To illustrate the effective performance of RPC, applications on real world datasets are carried out in Section 7, prior to summary of this paper in Section 8.

2 PRELIMINARIES
2.1 Principal Curves
The principal curve is the nonlinear extension of the first PCA in the sense of depicting a data skeleton. Let a dataset \(X = (x_1, x_2, \ldots, x_n), x_i \in \mathbb{R}^d\). The first PCA summarizes the data with a straight line, \(x = \mu + sw + \varepsilon\) where \(\mu\) is the mean of \(X, w\) is the line direction, \(s\) is the first component score and \(\varepsilon\) is the residual which is taken as noise in practice. Instead of a line, a principal curve summarizes the data with a smooth curve

\[ x = f(s) + \varepsilon \]

where \(f(s) = (f_1(s), f_2(s), \ldots, f_d(s)) \in \mathbb{R}^d\) and \(s \in \mathbb{R}\). The principal curve \(f\) was originally defined by Hastie and Stuetzle [26] as a smooth \((C^n)\) unit-speed \(\|f''\| = 1\) one-dimensional manifold in \(\mathbb{R}^d\) satisfying the self-consistence condition

\[ f(s) = E(x|s_t(x) = s), \]

where \(s = s_t(x) \in \mathbb{R}\) is the largest value so that \(f(s)\) has the minimum distance from \(x\). Mathematically, \(s_t(x)\) is formulated as [26]

\[ s_t(x) = \sup \{s : \|x - f(s)\| = \inf_s \|x - f(s)\|\}. \]

A curve \(f : \mathbb{R} \rightarrow \mathbb{R}^d\) is called a principal curve if it minimizes the expected squared distance between \(x\) and \(f\) which is denoted by [27]

\[ J(f) = E(\inf_s \|x - f(s)\|^2) = E\|x - f(s_t(x))\|^2. \]

Following Hastie and Stuetzle [26], researchers afterwards have proposed a variety of principal curve models to perform different tasks [27], [28], [29]. Some of them tried to first approximate the principal curve with a polyline [27] and then smooth it to meet the requirement of smoothness [26]. Some of them tried to find some key points and connect them smoothly to form a principal curve [31]. Therefore, expressions of principal curves are not explicit and tend to be a “black-box” which is hard to interpret. Some definitions of principal curves [30], [32] employed Gaussian mixture model to parametrically formulate the principal curve resulting in model bias. When the principal curve is used to perform a ranking task [28], it is required to be a “white-box” for easy interpretability and to meet the basic requirements of ranking, such as order-preserving.

2.2 Bézier Curves
A Bézier curve is a parametric model which is frequently used in computer graphics to model a smooth curve. The explicit formulation of the Bézier curve of degree \(k\) is given [45] parametrically by

\[ f(s) = \sum_{r=0}^{k} B^k_r(s)p_r, s \in [0, 1] \]

in terms of Bernstein polynomials

\[ B^k_r(s) = \binom{k}{r}(1-s)^{k-r}s^r, \]

\[ \binom{k}{r} = \frac{k!}{r!(k-r)!}. \]

In Eq. (4), \(p_r \in \mathbb{R}^d\) are control points of the Bézier curve. Taking \(p_0\) and \(p_k\) to be end points, the Bézier curve is regarded as the nonlinear interpolation of end points.

The Bézier curve has some attractive geometric properties in practice [45], four of which are listed as follows:

- **Affine Compatibility.** The Bézier construction is compatible with affine transformations. Let \(L\) be an affine map. \(L\) might perform scaling, translation and rotation. The affine transformation on the Bézier curve is directly imposed on the control points

\[ L(f(s)) = \sum_{r=0}^{k} B^k_r(s)L(p_r). \]

- **Invariance under Interval Affine Transformation.** The Bézier curve is invariant to interval affine from \([0, 1]\) to \([a, b]\), which reads algebraically

\[ f(s) = f\left(\frac{u-a}{b-a}\right) = \sum_{r=0}^{k} B^k_r\left(\frac{u-a}{b-a}\right)p_r, u \in [a, b]. \]

- **Convex Hull.** For \(0 \leq s \leq 1\), the Bézier curve lies entirely in the convex hull of its control points.

- **Reduced-Order Derivatives.** The Bézier curve is differentiable at any point in \([0, 1]\). Its derivative is a reduced-order Bézier curve with the form of

\[ f'(s) = \sum_{r=0}^{k-1} B^{k-1}_r(s)(p_{r+1} - p_r) \]

which is also differentiable. Therefore, a Bézier curve of degree \(k\) has derivatives of all orders. Quadratic and cubic Bézier curves are most common in practice, where \(k = 2\) and \(3\) respectively. Particularly for the cubic Bézier curve, Eq. (4) has the matrix form of

\[ f(s) = PMz. \]

where

\[ P = (p_0, p_1, p_2, p_3) \]
Hu et al. [35] proved that in two-dimensional space a cubic Bézier curve with constraints on control points is capable of depicting four basic nonlinear monotonicity shapes besides the straight line (illustrated in Fig. 3), while the quadratic Bézier curve fails to give a "S"-type curve which limits its application in ranking. Higher degree Bézier curves are more computationally expensive. When more complex shapes are needed, low order Bézier curves can be patched together [45].

### 3 Backgrounds

Consider ranking a set of \( n \) objects \( \mathcal{S} = \{a_1, a_2, \ldots, a_n\} \) according to \( d \) real-valued attributes (or indicators, features) \( V = \{v_1, v_2, \ldots, v_d\} \). Numerical observations of one object \( a \in \mathcal{S} \) on all the attributes comprise an item which is denoted as a vector \( x \) in \( d \)-dimensional space \( \mathbb{R}^d \). Ranking objects in \( \mathcal{S} \) is equivalent to ranking data points \( X = \{x_1, x_2, \ldots, x_n\} \). That is, to give the ordering of \( a_{i_1} \preceq a_{i_2} \preceq \cdots \preceq a_{i_n} \) can be achieved by discovering the ordering of \( x_{i_1} \preceq x_{i_2} \preceq \cdots \preceq x_{i_n} \) where \( \{i_1, i_2, \ldots, i_n\} \) is a permutation of \( \{1, 2, \ldots, n\} \) and \( x_i \preceq x_j \) means that \( x_i \) precedes \( x_j \). As there is no label to help with ranking, it is an unsupervised ranking problem from multi-attribute data.

Mathematically, ranking task is to provide a list of totally ordered points. A total order is a special partial order which requires comparability in addition to the requirements of reflexivity, antisymmetry and transitivity for the partial order [40]. Let \( x \) and \( y \) are one pair of points in \( X \). If \( x \) and \( y \) are different, they have the ordinal relation of either \( x \preceq y \) or \( y \preceq x \). If \( x \preceq y \) and \( y \preceq x \), then \( y = x \) which infers that \( x \) and \( y \) are the same thing. For unsupervised ranking, we defined the total order by Definition 1 associated with the self-dual proper cone \( \mathbb{R}_+^d = \{\rho : \rho^T x \geq 0, \forall x \in \mathbb{R}_+^d\} \) [40].

**Definition 1.** The order between two vectors \( x \) and \( y \) in the subset \( \mathcal{S} \subseteq \mathbb{R}^d \) is defined to be

\[
x \preceq y \iff (\delta_1(y_1 - x_1), \delta_2(y_2 - x_2), \ldots, \delta_d(y_d - x_d)) \in \mathbb{R}_+^d
\]

where \( x = (x_1, x_2, \ldots, x_d)^T \), \( y = (y_1, y_2, \ldots, y_d)^T \),

\[
\delta_j = \begin{cases} 1, & j \in \mathbf{E} \\ -1, & j \in \mathbf{F} \end{cases}
\]

\( \mathbf{E} \cap \mathbf{F} = \emptyset \) and \( \mathbf{E} \cup \mathbf{F} = \{1, 2, \ldots, d\} \).

For a given ranking task, \( x \) precedes \( y \) for \( x_j < y_j (j \in \mathbf{E}) \) and \( x_j \geq y_j (j \in \mathbf{F}) \). Let

\[
\alpha = (\delta_1, \delta_2, \ldots, \delta_d)^T
\]

and name it monotonicity index. For example, \( \alpha = (1, 1, -1, -1)^T \) for a ranking problem, \( \mathbf{E} = \{1, 2\} \) and \( \mathbf{F} = \{3, 4\} \) accordingly. For any pair of ranking candidates \( x \) and \( y \), \( x_j < y_j \) holds for \( j = 1, 2 \) while \( x_j > y_j \) holds for \( j = 3, 4 \). Therefore, the full ranking list \( x_{i_1} \preceq x_{i_2} \preceq \cdots \preceq x_{i_n} \) is in the same order of the increasing orders of the first and second attributes observations while also in the same order of the decreasing orders of the third and fourth attributes observations. \( \alpha \) is unique for one given ranking task and varies from task to task. It can be the prior knowledge along with attributes, or can be learned from data points \( X \).

As \( R \) is totally ordered, we prefer to grade each point with a real value to help with ranking. Assume \( \varphi : \mathbb{R}^d \mapsto R \) is the ranking function to assign \( x \) a score which provides the ordering of \( x \). \( \varphi \) is required to be order-preserving so that \( \varphi(x) \) has the same ordering in \( R \) as \( x \) in \( \mathbb{R}^d \). In order theory, an order-preserving function is also called isotone or monotone [41].

**Definition 2 ([41]).** A function \( \varphi : \mathbb{R}^d \mapsto R \) is called monotone (or, alternatively, order-preserving) if

\[
x \preceq y \implies \varphi(x) \leq \varphi(y)
\]

and strictly monotone if

\[
x \preceq y, x \neq y \implies \varphi(x) < \varphi(y)
\]

Order-preserving is the basic requirement for a ranking function. For a partially ordered set, \( \varphi \) should assign a score to \( x \) no more than the score to \( y \) if \( x \preceq y \). Moreover, if \( x \not\preceq y \) also holds, the score assigned to \( x \) must be smaller than the score to \( y \). As \( \mathcal{S} \) is totally ordered and different points should be assigned with different scores, the ranking function is required to be strictly monotone as stated by Eq. (17).

**Example 2.** In addition to the two indicators in Example 1, another two indicators are taken to evaluate life qualities of countries: IMR \(^3\) and Tub. \(^4\) It is easily known that the

\(^3\) Infant Mortality Rate per 1000 born.

\(^4\) New cases of infectious Tuberculosis per 100,000 of population.
life quality of one country would be higher if it has a higher LEB and GDP while a lower IMR and Tub. Let numerical observations on four countries to be \( x_i = (2.1, 62.7, 75.59, x_{M} = (11.3, 75.5, 12.30), x_{Q} = (32.1, 79.2, 6.4), \) and \( x_N = (47.6, 80.1, 3.3) \) respectively. By Eq. (13), they have the ordering \( x_1 \preceq x_M \preceq x_Q \preceq x_N \) with monotonicity index \( \alpha = (1.1, -1, -1)^T \), where \( E = \{1, 2\} \) and \( F = \{3, 4\} \). \( \alpha \) also discovers the relationship between indicators for ranking. GDP is in the same direction with LEB, but in the opposite direction with IMR and Tub. Let \( \phi(x_1) = 0.407, \phi(x_2) = 0.593, \phi(x_3) = 0.785 \) and \( \phi(x_4) = 0.891 \). Then \( \phi \) is a strictly monotone mapping which strictly preserves the ordering in \( \mathbb{R}^4 \).

**Definition 3 ([42]).** For a non-negative integer \( h, \phi(x) \) is of the differentiability class \( C^h \) if all the partial derivatives \( \frac{\partial^k \phi}{\partial x_1^{i_1} \partial x_2^{i_2} \cdots \partial x_d^{i_d}} \) exist and are continuous for all \( x \) in \( \text{dom} \phi \), where \( k \) is a non-negative integer, \( i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, d\} \) and \( h_1, h_2, \ldots, h_k \in \{0, 1, \ldots, h\} \) are integer sequences, and \( h_1 + h_2 + \cdots + h_k = h \).

**Theorem 1 ([40]).** Let \( \phi: \mathbb{R}^d \rightarrow \mathbb{R} \) be of the differentiability class \( C^h (h \geq 1) \), \( \phi \) is monotone if and only if
\[
\nabla \phi(x) \geq 0
\]
where \( 0 \) is the zero vector. \( \phi \) is strictly monotone if
\[
\nabla \phi(x) > 0
\]

Theorem 1 provides first-order conditions for monotonicity. Note that ‘>’ denotes a strict partial order [40]. Let
\[
\nabla \phi(x) = \left( \frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \ldots, \frac{\partial \phi}{\partial x_d} \right)^T.
\]
\[
\nabla \phi(x) > 0 \text{ infers } \frac{\partial \phi}{\partial x_j} > 0 \text{ for } j \in E \text{ and } \frac{\partial \phi}{\partial x_j} < 0 \text{ for } j \in F. \nabla \phi(x) > 0 \text{ infers that each component of } \nabla \phi(x) \text{ does not equal to zero. By the case of strict monotonicity in Theorem 1, } \nabla \phi(x) > 0 \text{ infers not only that } \phi \text{ is strictly monotone from } \mathbb{R}^d \text{ to } \mathbb{R}, \text{ but also that the value } s = \phi(x) \text{ is increasing with respect to } x_j(j \in E) \text{ and decreasing with respect to } x_j(j \in F). \text{ Vice versa, if } \frac{\partial \phi}{\partial x_j} \text{ is bigger than zero for } j \in E \text{ and smaller than zero for } j \in F, \text{ if } \nabla \phi(x) > 0 \text{ holds and infers } \phi \text{ is a strictly monotone mapping. Lemma 1 can be concluded immediately.}

**Lemma 1.** \( s = \phi(x) \) is strictly monotone if and only if \( s \) is strictly monotone along \( x_i \) with fixed the others \( x_j(j \neq i) \).

Further more, a strictly monotone mapping infers a one-to-one mapping that for a value \( s \in \text{rang}(\phi) \) there is exactly one point \( x \in \text{dom}(\phi) \) such that \( \phi(x) = s \). If the point \( x \) is denoted by \( x = f(s), f: \mathbb{R} \rightarrow \mathbb{R}^d \) is called the inverse mapping of \( \phi \) and inherits the property of strict monotonicity of its origin \( \phi \).

**Theorem 2.** Assume \( \nabla \phi(x) > 0 \). There exists an inverse mapping denoted by \( f: \text{rang}(\phi) \rightarrow \text{dom}(\phi) \) such that \( \nabla f(s) > 0 \) holds for all \( s \in \text{rang}(\phi) \), that is for \( \forall s_1, s_2 \in \text{rang}(\phi) \)
\[
s_1 < s_2 \implies f(s_1) \leq f(s_2), f(s_1) \neq f(s_2).
\]

Proof of Theorem 2 can be found in Appendix A, which can be found on the Computer Society Digital Library at [http://doi.ieeecomputersociety.org/10.1109/TKDE.2015.2441692](http://doi.ieeecomputersociety.org/10.1109/TKDE.2015.2441692). The theorem also holds in the other direction. Assuming \( f: \mathbb{R} \rightarrow \mathbb{R}^d \), if \( \nabla f(s) > 0 \), there exists an inverse mapping \( \phi: \text{rang}(f) \rightarrow \text{dom}(f) \) and \( \nabla \phi(x) > 0 \) holds for all \( x \in \text{rang}(f) \). Because of the one-to-one correspondence, \( f \) and \( \phi \) share the same geometric properties such as scale and translation invariance, smoothness and strict monotonicity [42].

### 4 Meta-Rules

Motivated by PageRank [9], we draw five essential items from ranking knowledge, named meta-rules, to serve as high-level assessments for unsupervised ranking results. These meta-rules can also guide the design of ranking functions. Assume a ranking function \( \phi: \mathbb{R}^d \rightarrow \mathbb{R} \), where \( \phi(x) \) outputs a real value \( s = \phi(x) \) as the ranking score for a given point \( x \). Five essential meta-rules are formalized to be scale and translation invariance, strict monotonicity, compatibility of linearity and nonlinearity, smoothness, and explicitness of parameter size.

#### 4.1 Scale and Translation Invariance

**Definition 4 ([43]).** A ranking rule is invariant to scale and translation if for \( x \preceq y \)
\[
\phi(x) \leq \phi(y) \iff \phi(L(x)) \leq \phi(L(y)).
\]
where \( L(\cdot) \) performs scale and translation.

This is a basic requirement for a ranking function and has been referred in multiple literatures (e.g., [7], [43]). In general, numerical observations on different indicators are taken on different quantity dimensions. In Example 1, GDP is measured in thousands of dollars while LEB ranges from 40 to 90 years. They are not in the same quantity dimensions. Preprocessing of datasets needs to be done according to normalization or standardization, both of which takes the transformation of scale and translation, while original order is required to be preserved. If let \( L \) be a linear transformation on \( \mathbb{R}^d \), we have \( x \preceq y \iff L(x) \preceq L(y) \) for \( x, y \in \mathbb{R}^d \) [43]. Therefore, a ranking function \( \phi(x) \) should produce the same ranking list before and after transformation of scaling and translating.

#### 4.2 Strict Monotonicity

**Definition 5 ([41]).** \( \phi(x) \) is strictly monotone if \( \phi(x_i) < \phi(x_j) \) for \( x_i \preceq x_j \) and \( x_i \neq x_j(i \neq j) \).

Monotonicity in Definition 2 is specified here as one of meta-rules for ranking. It has been widely used as ranking knowledge constraints in ordinal classification and regression [11], [36], [37]. Monotonicity helps improve these supervised tasks greatly by embedding it into learning models. For unsupervised ranking problems, ranking functions with monotonicity constraint are also expected to produce ranking lists closer to “ground-truth” than ranking functions without it. \( \phi(x_i) = \phi(x_j) \) holds if and only if \( x_i = x_j(i \neq j) \). In Example 1 (Fig. 2), \( x_1 \preceq x_2 \) and \( x_i \neq x_j \) indicate that a higher score should be assigned to \( x_2 \) than \( x_1 \). And so do \( x_1 \) and \( x_2 \). Therefore, the ranking function \( \phi(x) \) is required to be a strictly monotone mapping. \( \phi \) in Example 2 is to the point referred here.
4.3 Compatibility of Linearity and Nonlinearity

Definition 6. \( \varphi(x) \) has the compatibility of linearity and nonlinearity if \( \varphi(x) \) is able to depict the relationship of both linearity and nonlinearity.

Taking the ranking task in Example 1 for illustration, one has no knowledge about the relationship between LEB and the score. The score might be a either linear or nonlinear function of LEB. It is the similar case for the relationship between GDP and the score. Therefore, \( t = \varphi(x) \) should embody both of the linear and nonlinear relationships between \( t \) and \( x \). For the ranking task in Example 1, the ranking function \( \varphi \) should be a linear function of LEB for fixed GDP if LEB is linear with \( t \). Meanwhile, \( \varphi \) should also be a nonlinear function of GDP for fixed LEB if GDP is nonlinear with \( t \). Otherwise, model-bias might occur.

4.4 Smoothness

Definition 7. \( \varphi(x) \) is smooth if \( \varphi(x) \) is \( C^h \) \((h \geq 1)\).

In mathematical analysis, a function is called smooth if it has derivatives of all orders (Definition 3). Yet a ranking function \( \varphi(x) \) is required to be of class \( C^h \) where \( h \geq 1 \). That is, \( \varphi(x) \) is required to be continuous and have at least the first-order derivative \( \nabla \varphi(x) \) at each point in its domain. For the requirement of continuity \( (C^0) \), the ranking function is not allowed a suddenly jump at some point such that minor differences of two objects produce a large gap of ranking scores. For the requirement of the first-order derivative \( (C^1) \), \( \nabla \varphi(x) \) guarantees that \( \varphi(x) \) will exert a consistent and acceptable ranking rule for all objects. Otherwise, it would bring a problem to ranking, which is illustrated in Fig. 4. The polyline is of class \( C^0 \) but not of class \( C^1 \) because it is continuous but not differentiable at the connecting vertex of the two lines. GDP abruptly changes its impact on the ranking score at the non-differentiable point, resulting in that the same score is assigned to two apparently ordered points which are projected to the same connecting vertex (Example 3).

Example 3. In Fig. 4, for points \( x_7 = (75, 8) \) and \( x_8 = (76, 14) \), \( x_7 \) should be ranked higher than \( x_8 \) because of its higher LEB and higher GDP. However, they are projected to the same vertex \((74.5, 15)\) and will be assigned the same score such that they can not be distinguished by the ranking rule of the piecewise line.

4.5 Explicitness of Parameter Size

Definition 8. \( \varphi(x) \) has the property of explicitness if \( \varphi(x) \) has known parameter size for a fair comparison among ranking models.

Generally, nonlinear regression will introduce the model selection problem. However, conflicting results would occur with conventional methods of model selection. For example, Bayesian information criterion (BIC) for model selection is given by [25]

\[
BIC = Error + \text{ModelSize}
\]

and there are two models A and B with:

\[
BIC_A = 10 = 2 + 8
\]

\[
BIC_B = 9 = 6 + 3.
\]

With the BIC criterion, the model B is preferred, in conflict with the result with error criterion that the model A is preferred. To avoid this conflict, the explicitness of parameter size ensures a fair comparison between models of the same parameter size.

5 Ranking Principal Curves

For unsupervised ranking, we present a RPC model which follows all the five meta-rules. The RPC is parametrically designed with a cubic Bézier curve of strict monotonicity.

5.1 RPC Model

Example 4. As we know, the life quality of one country would be higher if it has a higher LEB and GDP. Let numerical observations of two countries be \( x_T = (78, 12) \) and \( x_P = (72, 21) \). One would get the order of \( x_P \leq x_T \) with LEB and \( x_T \leq x_P \) with GDP. It is a very often phenomenon that two objects get contrary orders on different indicators.

In real applications, the order of precedence cannot be directly determined according to comparison dimension by dimension due to the fact that multi-attribute objects would show different ranking orders for different indicators. In modeling, we assume that the contrary was incurred by noise in numerical observations of objects. That is, \( x = x_{true} + \epsilon \), where \( \epsilon \) is the noise to \( x_{true} \) which is the noiseless observation and \( x \) is the observed data. The score \( s \) should be produce for \( x_{true} = x - \epsilon \) by a ranking function, that is \( s = \varphi(x - \epsilon, \theta) \). As a ranking function, \( \varphi \) is assumed to be strictly monotone. Thus by Theorem 2, there exists an inverse function \( f \) for \( \varphi \) such that

\[
x = f(s, \theta) + \epsilon
\]

which is the principal curve model [26]. The inverse function can be taken as the generating function for numerical observations of ranking candidates from the score \( s \) which can be regarded to be pre-existing. The principal curve is used to approximate the latent generating function. The design of a ranking function turns to find the corresponding principal curve for the given numerical observations.

Correspondingly, five meta-rules for ranking functions \( \varphi \) should be also followed by the principal curve \( f \) for ranking. However, general principal curve models are probably not suitable to carry out ranking tasks. As the degeneration of the principal curve, the first PCA [17] can well depict the ranking skeletons of datasets like ellipses but fail for the cases of crescents (e.g., Fig. 5a). Polyline approximations [27] of the principal curve might go against smoothness and
strict monotonicity (e.g., Fig. 5b). A smooth principal curve model (e.g., [28]) would still go against strict monotonicity (e.g., Fig. 5c). Going against smoothness and strict monotonicity might bring problems which are illustrated in Example 1 (Fig. 2) and Example 3 (Fig. 4). Within the framework of Fig. 1, all the five meta-rules can be modeled as constraints to the ranking function. Since a principal curve is defined to be smooth and invariant to scale and translation [26], the constraint of strict monotonicity would make it be capable of performing ranking tasks (e.g., Fig. 5d). Naturally, the principal curve should have a known parameter size for a fair comparison. We present Definition 9 for unsupervised ranking with a principal curve.

Definition 9. A one-dimensional curve \( f(s, \theta) \) in \( d \)-dimensional space is called a RPC if \( f(s, \theta) \) is a strictly monotone principal curve of given data cloud and it is explicitly expressed with known parameters \( \theta \) of limited size.

5.2 RPC Formulation with Bézier Curves

Remembering attractive properties of the cubic Bézier curve (Section 2.2), we select the cubic Bézier curve with constraints on control points to parametrically model an RPC. Specifically, the RPC is modeled to be

\[
\mathbf{f}(s) = \sum_{i=0}^{3} B_i^3(s) \mathbf{p}_i, \quad s \in [0, 1],
\]

where control points are RPC parameters which is of size \( 4d \). Moreover, we specify four propositions for the cubic Bézier curve following the other four meta-rules.

Proposition 1. A cubic Bézier curve is invariant to scale and translation.

Proof. Specifying the affine transformation \( \mathbf{L} \) to be scale and translation, \( \mathbf{L} \) can be imposed directly on control points without changing the score \( s \). That is

\[
\mathbf{L} \mathbf{f}(s) + \mathbf{b} = \mathbf{LPMz} + \mathbf{b} = (\mathbf{AP} + \mathbf{b}) \mathbf{Mz},
\]

where \( \mathbf{A} \) is a diagonal matrix with scaling factors to dimensions and \( \mathbf{b} \) is the translation vector.

This proposition allows data preprocessing of regularization without changing ranking scores. In this work, we regularize data into \([0, 1]^d\).

Proposition 2. \( f(s) \) is strictly monotone for \( s \in [0, 1] \) with \( \mathbf{p}_0 = \frac{1}{2}(1 - \alpha), \mathbf{p}_3 = \frac{1}{2}(1 + \alpha) \) and \( \mathbf{p}_1, \mathbf{p}_2 \in (0, 1)^d \).

Proof. Let end points are denoted by \( \mathbf{p}_0 = \frac{1}{2}(1 - \alpha) \) and \( \mathbf{p}_3 = \frac{1}{2}(1 + \alpha) \). Control points \( \mathbf{p}_1 \) and \( \mathbf{p}_2 \) are the determinants for monotonic shapes of the cubic Bézier curves (Fig. 3). It has been proved [35] that \( \nabla f(s) \geq 0 \) for \( \mathbf{p}_1, \mathbf{p}_2 \in [0, 1]^d \) and \( \nabla f(s) = 0 \) occurs in the border of \([0, 1]^d\). When we restrict \( \mathbf{p}_1 \) and \( \mathbf{p}_2 \) in the interior of \([0, 1]^d\), that is \((0, 1)^d\), we get \( \nabla f(s) \neq 0 \) which infers strict monotonicity by Theorem 1.

Proposition 3. A cubic Bézier curve has the capabilities of linearity and nonlinearity.

Proof. Eq. (27) is a nonlinear function of \( s \) due to \( B_i^3(s) \) (Eq. (5)). If \( \mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \) are evenly spaced (i.e., form an arithmetic progression), the cubic Bézier curve becomes a linear function \( f(s) = \mathbf{p}_0 + s(\mathbf{p}_1 - \mathbf{p}_0) \) which runs from \( \mathbf{p}_0 \) to \( \mathbf{p}_3 \) as \( s \) runs from 0 to 1.

Proposition 4. A cubic Bézier curve is of the differentiability class \( C^4 \).

Proposition 4 is a directly results of the property of reduced-order derivatives in Section 2.2. What is the most important, there always exists an RPC parameterized by a cubic Bézier curve of strict monotonicity for a group of numerical observations. The existence has failed to be proved in many principal curve models [26], [28], [31].

Theorem 3. Assume that \( \mathbf{x} \) is the numerical observation of a ranking candidate and that \( E\|\mathbf{x}\|^2 < \infty \). There exists \( \mathbf{P}^* \in [0, 1]^d \) such that \( f(\mathbf{s}) = \mathbf{P}^* \mathbf{Mz} \) is strictly monotone and

\[
J(\mathbf{P}^*) = \inf \left\{ J(\mathbf{P}) = E\left( \inf_{\mathbf{s}} \|\mathbf{x} - \mathbf{PMz}\|^2 \right) \right\}.
\]

Proof of Theorem 3 can be found in Appendix B, available online.

6 RPC LEARNING ALGORITHM

To perform unsupervised ranking from the multi-attribute numerical observations of ranking candidates \( \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \), an RPC modeled by a cubic Bézier curve should be learned to achieve the infimum of the estimation of \( J(\mathbf{P}) \) in Eq. (29). By the principal curve definition proposed by Hastie and Stuetze[26], the RPC is the curve which minimizes the summed residual \( \epsilon \). Therefore, the ranking task is formulated as a nonlinear optimization problem

\[
\min_{\mathbf{P}, \mathbf{s}} J(\mathbf{P}, \mathbf{s}) = \frac{n}{\mathbf{s}} \|\mathbf{x}_i - \mathbf{PMz}_i\|^2
\]

s.t. \( \mathbf{P} \in [0, 1]^d \), \( \mathbf{s} \in [0, 1], \quad i = 1, 2, \ldots, n \).
where Eq. (31) determines $s_i$ to find the point on the curve which has the minimum residual to reconstruct $x_i$ by $f(s_i)$. Obviously, a local minimizer $(P^i, s^i)$ can be achieved in an alternating minimization way

$$P^{(t+1)} = \arg \min \sum_{i=1}^{n} ||x_i - PMZ_i^{(t)}||^2$$

$$\frac{\partial P^{(t+1)}Mz}{\partial s} \left( x_i - P^{(t+1)}Mz \right)_{s=s_i^{(t+1)}} = 0$$

where $t$ means the $t$th iteration.

The optimal solution of Eq. (32) has an explicit expression. Associate $X$ with $Z$

$$Z = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ s_1 & s_2 & \cdots & s_n \\ \bar{s}_1 & \bar{s}_2 & \cdots & \bar{s}_n \end{pmatrix} = (z_1, z_2, \ldots, z_n)$$

and Eq. (30) can be rewritten in matrix form

$$J(P, s) = ||X - PMZ||_F$$

$$= tr(X^T X) - 2tr(PMZ^T X^T Z)$$

$$+ tr(PMZ^T M^T P^T).$$

Setting the derivative of $J$ with respect to $P$ to zero

$$\frac{\partial J}{\partial P} = 2\left(P(MZ)(MZ)^T - X(MZ)^T \right) = 0$$

and remembering $A^+ = A^T(AA^T)^+$, we get an explicitly expression for the minimum point of Eq. (30)

$$P = X(MZ)^T \left((MZ)(MZ)^T \right)^+ = X(MZ)^+$$

where $(\cdot)^+$ takes pseudo-inverse computation. Based on the $t$th iterative results $Z^{(t)}$, the optimal solution can be given by substituting $Z^{(t)}$ into Eq. (37) which is $P^{(t+1)} = X(MZ)^{+}$. However, $(MZ)^{+}$ is computationally expensive in numerical experiments and $X$ is always an ill-conditioned matrix which has a high condition number, resulting in that a very small change in $Z^{(t)}$ would produce a tremendous change in $P^{(t+1)}$, $Z^{(t)}$ is not the optimal solution of Eq. (30) but an intermediate result of the iteration, and $P^{(t+1)}$ would thereby go even far away from the optimal solution. To settle out the problem, we employ the Richardson iteration [38] with a preconditioner $D$ which is a diagonal matrix with the $L_2$ norm of columns of $(MZ)(MZ)^T$ (Eq. (37)) as its diagonal elements. Then $P^{(t+1)}$ is updated iteratively according to

$$P^{(t+1)} = \gamma^{(t)}(P^{(t)}(MZ)(MZ)^T) - X(MZ)^{+}D^{-1},$$

where $\gamma^{(t)}$ is a scalar parameter such that the sequence $P^{(t)}$ converges. In practice, we set

$$\gamma^{(t)} = \frac{2}{\lambda_{min}^{(t)} + \lambda_{max}^{(t)}}$$

where $\lambda_{min}^{(t)}$ and $\lambda_{max}^{(t)}$ is the minimum and maximum eigenvalues of $(MZ)(MZ)^T$ respectively [50]. After getting $P^{(t+1)}$, the score vector $s^{(t+1)}$ can be calculated as the solution to Eq. (33). Eq. (33) is a quintic polynomial equation which rarely has explicitly expressed roots.

**Algorithm 1.** Algorithm to learn an RPC.

**Input:**

- $X$: data matrix;
- $\xi$: a small positive value;

**Output:**

- $P^*$: control points of the learned Bézier curve
- $s^*$: the score vector of objects in the set.

1. Normalize $X$ into $[0, 1]^d$;
2. Initialize $P^{(0)}$;
3. while $\Delta J > \xi$ do
4. Adopt GSS to find the approximate solution $s^{(t)}$;
5. Compute $P^{(t+1)}$ using a preconditioner;
6. if $\Delta J < 0$ then
7. break;
8. end if
9. end while

Algorithm 1 summarizes the alternative optimization procedure. Before performing the ranking task, numerical observations of objects should be normalized into $[0, 1]^d$ by

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}},$$

where $\hat{x}$ is the normalized vector of $x$, $x_{min}$ the minimum vector and $x_{max}$ the maximum vector. Grading scores would be unchanged as scaling and translating are only performed on control points and end points (Eq. (28)) without changing the interpolation values. In Step 2, we initialize the end points as $p_0 = \frac{1}{2}(1 - \alpha)$ and $p_3 = \frac{1}{2}(1 + \alpha)$, and randomly select samples as control points. During learning procedure, $P^{(t)}$ is automatically learned to make a Bézier curve be an RPC. In Step 6, $\Delta J < 0$ occurs when $J$ begins to increase. In this case, the algorithm stops updating $(P^{(t)}, s^{(t)})$ and gets a local minimum $J$. Proposition 5 guarantees the convexity of the sequence found by RPC learning algorithm (proof can be found in Appendix C, available online). Therefore, the RPC learning algorithm finds a converging sequence of $(P^{(t)}, s^{(t)})$ to achieve the infimum in Eq. (29).

**Proposition 5.** If $P^{(t)} \rightarrow P^*$ as $t \rightarrow \infty$, $J(P^{(t)}, s^{(t)})$ is a decaying sequence which converges to $J(P^*, s^*)$ as $t \rightarrow \infty$.

Algorithm 1 converges in limited steps. In each step, $P$ is updated in $4 \times d$ size and scores for points are calculated in $n$ size. When iteration stops, ranking scores are produced along with $P$. In summary, the computational complexity of RPC unsupervised ranking model is $O(n)$.

RPC learning algorithm learns a ranking function in a completely different way from traditional ranking methods. The ranking function is in constraints of five ranking meta-rules. Integrating meta-rules with ranking functions makes
the ranking rule be more in line with human knowledge about ranking problems. As a high level knowledge, these meta-rules are capable of evaluating ranking performance. Moreover, ranking is carried out following the principle of unsupervised ranking, "let the data speak for themselves". For unsupervised ranking, the structure of the dataset contains the ordinal information between objects. Ideally, the RPC can thread through all the objects successively. Yet in practice, the most influential indicators are selected to estimate the order of objects and noise exists. In the case we know nothing about the rest factors, we would better to minimize the effect which we formulate to be error $\varepsilon$. Therefore, minimizing errors is adopted as the learning objective.

7 EXPERIMENTS

Since no ground-truth ranking labels are available, quantitative evaluations can be hardly carried out. In this part, RPC is compared with two heuristic rank aggregation methods and two unsupervised ranking methods on five meta-rules in the sense of meta-rules and artificial data. Two real applications, world university ranking and journal ranking, are carried out to validate RPC’s applicability.

7.1 Comparisons on Meta-Rules

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_1$ Value</td>
<td>$x_2$ Value</td>
<td>Score Order</td>
<td>Score Order</td>
<td>Score Order</td>
<td>Score Order</td>
</tr>
<tr>
<td>$A$</td>
<td>0.30</td>
<td>0.25</td>
<td>3</td>
<td>2</td>
<td>0.55</td>
<td>3</td>
</tr>
<tr>
<td>$B$</td>
<td>0.25</td>
<td>0.35</td>
<td>2</td>
<td>2</td>
<td>0.80</td>
<td>2</td>
</tr>
<tr>
<td>$C$</td>
<td>0.70</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
<td>1.60</td>
<td>1</td>
</tr>
</tbody>
</table>

1 MPM ranked lists are in their inverse orders (MPM Package is available on http://www.cs.toronto.edu/~mvonkovs/).

The ending of the text is missing. Context is provided as follows:

<table>
<thead>
<tr>
<th>Meta-Rule</th>
<th>BC</th>
<th>CombSUM</th>
<th>MPM</th>
<th>Elmap</th>
<th>RPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale and Translation Invariance</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Strict Monotonicity</td>
<td>Y</td>
<td>Y</td>
<td>U</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Capacities of Linearity and Nonlinearity</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Smoothness</td>
<td>N</td>
<td>$C^1$</td>
<td>U</td>
<td>$C^3$</td>
<td>$C^4$</td>
</tr>
<tr>
<td>Explicitness of Parameter Size</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 3: Meta-Rules Followed by Different Ranking Methods (Y - Yes, N - No, U - Unknown)

<table>
<thead>
<tr>
<th>Complexity</th>
<th>$O(dn)$</th>
<th>$O(n^{5/3})$</th>
<th>$O(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Size</td>
<td>$n$</td>
<td>$k(2 + d)$</td>
<td>$4d$</td>
</tr>
</tbody>
</table>

Table 4: Comparisons of Complexities and Parameter Sizes

For simple aggregation methods, we select Borda count (BC) [22] and CombSUM [46] for comparisons. As an order-based method, BC cannot efficiently detect ordinal information embedded in the numerical observations due to its non-smoothness, illustrated in Table 2. BC provides the same order when $A$ in Table 2(a) is changed to $A'$ in Table 2(b). Furthermore, BC implies nonlinear relationships between indicators and ranking scores, resulting in a model bias. As a score-based method, CombSUM provides different scores for $A$ and $A'$ but still the same ranking lists. Whether score-based or order-based, the simple aggregation methods apply the same aggregation strategy for all cases and do not consider the variability of different ranking tasks.

Among several unsupervised ranking aggregation methods, MPM is the latest developed method [21]. It shows very good performance on aggregating meta-search results in the sense of NDCG and MAP on LETOR [13]. MPM has an implicit ($s = \psi(x)$, $\psi$ is implicit) ranking function and can deal with partial lists such that MPM has wide applications on supervised and unsupervised ranking. In contrast, RPC developed recently, most of which focus on the information retrieval of web search [3], [19], [21], [22], [23]. Ranking for web search is query-oriented where websites are ranked by some strategies with the referenced query [3]. For ranking tasks, some researchers prefer to aggregate different ranking lists of the same set of objects in order to get a consensus order. We compare RPC with two heuristic rank aggregation methods and two unsupervised ranking methods and summarize comparisons on meta-rules in Table 3. Computation complexities and parameter sizes are summarized in Table 4.
is an explicit ranking function and requires full lists of attributes such that RPC is unsuitable for aggregating search results of different engines (like LETOR). However, the competitive advantage of RPC lies in visualization such that the user can have a clear scope of the “ranking coordinate”. The parameterized RPC makes it easier to interpret the ranking rules for ranking candidates.

7.1.2 Comparison with Elmap

RPC is motivated from Elmap [28] yet makes the principal curve application in ranking more systematical. Compared to Elmap, RPC takes the advantages of smoothness and strict monotonicity which have been widely used in ordinal regression and classification. Figs. 2 and 4 show the necessities of these meta-rules for principal curves to perform ranking tasks. For the synthetical data in Table 2, Elmap produces the same score for \( A \) and \( B \) due to the non-smoothness (with Elmap Package available on http://www.ihes.fr/~zinovyev/vida/elmap/index.htm) although it also detects the numerical information. They are compared also in the sense of fitting MSE as principal curves for numerical observations of ranking candidates (Fig. 6). As Elmap produces different principal curves for different nodes number assignments, we compare two cases of nodes number settings: Elmap with 4 points (Elmap\( _4 \), the same parameter numbers with RPC), and Elmap with auto selected points (Elmap\( _{Auto} \)). In the sense of fitting MSE, RPC obtains lower residual than both Elmap\( _4 \) and Elmap\( _{Auto} \).

7.2 Applications

Unsupervised ranking of multi-attribute objects has a widely applications. Taking the journal ranking task for illustration, there have been many indices to rank journals, such as impact factor (IF) [51] and Eigenfactor [52]. Different indices reflect different aspects of journals and provide different ranking lists for journals. Different aggregating methods may produce similar ranking results, but the ranking list of the method which matches the prior information the most will be preferred by users. RPC model is proposed as a new ranking framework following the prior information which provides an ordering along the “ranking skeleton” of data distribution.

In this paper, we perform ranking tasks with RPCs to produce comprehensive evaluations on two open access datasets of countries and journals with the open source software Scilab (5.4.1 version) on a Ubuntu 12.04 system with 4 GB memory. Due to space limitation, we just list parts of their ranking lists. The whole lists are available in the supplemental materials. We also carry out ranking on universities, the ranking list of which is also available in the supplemental materials.

7.2.1 Results on Life Qualities of Countries

Zinovyev and Gorban [12] ranked 171 countries by life qualities of countries with data driven from GAPMINDER (http://www.gapminder.org/) based on four indicators as in Example 2. For comparison, we use the same four GAPMINDER indicators as in [12]. The RPC learned by Algorithm 1 is shown in two-dimensional visualizations in Fig. 7 and part of the ranking list is illustrated in Table 5.

RPC portrays the development trends, based on the noised observations, with different shapes, including linearity and nonlinearity (Fig. 7). \( \alpha = [1, 1, -1, -1]^T \) (just as Example 2) indicates the monotonicity directions for indicators. In the beginning, a small amount of GDP increasing brings about tremendous increasing of LEB and tremendous decreasing of IMR and Tub. When GDP exceeds $14300 (0.2 as normalized value in Fig. 7) per person, increasing GDP does result in little LEB increase, so does IMR and Tub decrease. As a matter of fact, it is hard to improve further LEB, IMR and Tub when they are close to the limit of human evolution. When control points (provided in the bottom of Table 5) have overlap values in one dimension (e.g., GDP), it means this dimension is a “C”-shape curve. Otherwise, the dimension is a “S”-shape curve. RPC scores of countries are nonlinear interpolation values between the best country (Luxembourg) and the worst country (Swaziland) with the control of \( p_1 \) and \( p_2 \) to determine their specific developing trends.
7.2.2 Results on Journal Ranking

We also apply RPC model to rank journals with data accessible from the Web of Knowledge (http://wokinfo.com/) which is affiliated to Thomson Reuters. Thomson Reuters publishes annually journal citation reports (JCR) which provide information about academic journals in the sciences and social sciences. JCR2012 reports citation information with indicators of impact factor, 5-year impact factor, Immediacy Index, Eigenfactor Score, and Article Influence Score. After journals with data missing are removed from the data table (58 out of 451), RPC model tries to provide a comprehensive ranking list of journals in the categories of computer science: artificial intelligence, cybernetics, information systems, interdisciplinary applications, software engineering, theory and methods. Table 6 gives part of the ranking list produced by RPC model based on JCR2012. Two-dimensional visualization of the RPC can be found in supplemental materials.

For this ranking task, a journal will rank higher with a higher value for each indicator, that is 

\[ \alpha = [1, 1, 1, 1]. \]

From Table 6, IEEE Transactions on Knowledge and Data Engineering (TKDE) is ranked in a higher place than IEEE Transactions on Systems, Man, and Cybernetics-Part A (SMCA) although SMCA has a higher IF (2.183) than TKDE (1.892). The lower influence score (0.767) of SMCA brings it down the ranking list (vs. 1.129 for TKDE). Therefore, TKDE gets a higher comprehensive evaluating score and wins a higher ranking place in the ranking list. This means that one indicator does not tell the whole story of ranking lists. RPC produces a ranking list for journals taking into account several indicators of different aspects.

7.3 Summary

Unsupervised ranking tasks aim at providing a ranking list for ranking candidates. Different ranking models may produce similar ranking results. It is hard to determine which result would be preferred by users. However, meta-rules are presented to evaluate whether the ranking models are “proper” or not, and the model following all the meta-rules is

### Table 6

<table>
<thead>
<tr>
<th>Title</th>
<th>Impact Factor (IF)</th>
<th>5-Year IF</th>
<th>Immediacy Index</th>
<th>Eigenfactor</th>
<th>Influence Score</th>
<th>RPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE T PATTERN ANAL</td>
<td>4.795</td>
<td>7</td>
<td>6.144</td>
<td>5</td>
<td>0.625</td>
<td>26</td>
</tr>
<tr>
<td>ENTERP INF SYST UK</td>
<td>9.256</td>
<td>1</td>
<td>4.771</td>
<td>10</td>
<td>2.682</td>
<td>2</td>
</tr>
<tr>
<td>J STAT SOFTW</td>
<td>4.910</td>
<td>4</td>
<td>5.907</td>
<td>6</td>
<td>0.753</td>
<td>18</td>
</tr>
<tr>
<td>MIS QUART</td>
<td>4.659</td>
<td>8</td>
<td>7.474</td>
<td>2</td>
<td>0.705</td>
<td>21</td>
</tr>
<tr>
<td>ACM COMPUT SURV</td>
<td>3.543</td>
<td>21</td>
<td>7.854</td>
<td>1</td>
<td>0.421</td>
<td>56</td>
</tr>
<tr>
<td>DECIS SUPPORT SYST</td>
<td>2.001</td>
<td>51</td>
<td>3.037</td>
<td>43</td>
<td>0.196</td>
<td>169</td>
</tr>
<tr>
<td>COMPUT STAT DATA AN</td>
<td>1.304</td>
<td>156</td>
<td>1.449</td>
<td>180</td>
<td>0.415</td>
<td>61</td>
</tr>
<tr>
<td>IEEE T KNOWL DATA EN</td>
<td>1.892</td>
<td>82</td>
<td>2.426</td>
<td>72</td>
<td>0.217</td>
<td>152</td>
</tr>
<tr>
<td>MACH LEARN</td>
<td>1.467</td>
<td>133</td>
<td>2.143</td>
<td>56</td>
<td>0.373</td>
<td>70</td>
</tr>
<tr>
<td>IEEE T SYST MAN CY A</td>
<td>2.183</td>
<td>53</td>
<td>2.44</td>
<td>68</td>
<td>0.465</td>
<td>46</td>
</tr>
</tbody>
</table>
the most preferred. For the models which satisfy all the meta-rules and need further evaluation, we will apply a fitting criterion, such as MSE (Fig. 6), to determine the final model.

8 Conclusions

Ranking and its tools have and will have an increasing impact on the behavior of human, either positively or negatively. However, those ranking activities are still facing many challenges which have greatly restrained to the rational design and utilization of ranking tools. Generally, ranking in practice is an unsupervised task which encounters a critical challenge that there is no ground truth to evaluate the provided lists.

Motivated by [9], [10], [11], this work presents five essential meta-rules (there might be other meta-rules to study on) as a high-level evaluation for unsupervised ranking performance. They are scale and translation invariance, strict monotonicity, linear/nonlinear compatibility, smoothness and explicitness of parameter size. They can also guide the design of ranking functions as constraints. Furthermore, we propose a RPC model which meets all the five meta-rules. Parametrically, RPC is modeled with a cubic Bézier curve by restricting control points in the interior of the hypercube \([0, 1]^d\), which can be learned from the data distribution without human interventions. Experiments on life qualities of countries and journals of computer sciences show that the proposed RPC model is applicable in practice.

However, RPC model requires full lists of attributes such that it cannot deal with missing value fields and partial lists. Considering that it is common in practice, we will endeavor on this problem in our future works. In application, RPC model is also expected to be carried out on more complex datasets, as well as on unsupervised feature selection [53] in the near future.

Acknowledgments

The authors appreciate very much the advice from the machine learning crew in NLPR. This work is supported in part by NSFC (No. 61273186) for C.-G. Li and B.-G. Hu, and NSFC (No. 61271430 and 61332017) for X. Mei.

References


Chun-Guo Li received the BSc and MSc degrees in mathematics from Hebei University, Baoding, China, in 2003 and 2006, respectively. She is currently working toward the PhD degree in computer science at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing, China. From 2008, she worked as a lecturer in Hebei University. Her research interests are in machine learning and data mining. She is a member of the IEEE.

Xing Mei received the PhD degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), in 2009. During March 2009–June 2011, he was a researcher at the Samsung Advanced Institute of Technology (SAIT), China Lab. He is currently an assistant researcher at NLPR, CASIA, and a research scholar at the University at Albany, State University of New York. His research interests lie in computer graphics and computer vision.

Bao-Gang Hu received the MSc degree from the University of Science and Technology, Beijing, China, in 1983, and the PhD degree from McMaster University, Canada, in 1993, all in mechanical engineering. From 1994 to 1997, he was a research engineer and senior research engineer at C-CORE, Memorial University of Newfoundland, Canada. Currently, he is a professor with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Science, Beijing, China. From 2000 to 2005, he was the Chinese director of LIAMA (the Chinese-French Joint Laboratory for Computer Science, Control, and Applied Mathematics). His main research interests include pattern recognition and plant growth modeling. He is a senior member of the IEEE.