

# A Rank Aggregation Algorithm for Performance Evaluation in Modern Sports Medicine with NMR-based Metabolomics\*

V. Vigneron<sup>a</sup> and H. Maaref<sup>b</sup>

Univ. Evry, Université Paris-Saclay, IBISC EA 4526, Evry, France

**Keywords:** Deep Learning, Pooling Function, Rank Aggregation, LBP, Segmentation, Contour Extraction.

**Abstract:** In most research studies, much of the gathered information is qualitative in nature. This article focuses on items for which there are multiple rankings that should be optimally combined. More specifically, it describes a supervised stochastic approach, driven by a Boltzmann machine capable of ranking elements related to each other by order of importance. Unlike classic statistical ranking techniques, the algorithm does not need a voting rule for decision-making. The experimental results indicate that the proposed model outperforms two reference rank aggregation algorithms, ELECTRE IV and VIKOR, and it behaves more stable when encountering noisy data.

## 1 INTRODUCTION AND RELATED WORKS

In the last decades, the field of multiple criteria decision-making (MCDM) has received considerable attention in engineering, sciences, and humanities as they are extremely efficient in situations where policymakers need to decide priorities (Yazdani et al., 2017).

There are optimal resolution procedures like linear programming or nonlinear optimization for solving problems governed by single criteria. But real-life situations demand the evaluation of a set of alternatives against multiple criteria and are typically structured as MCDM problems (Thakkar, 2021; Rahman et al., 2017). When a decision needs to be made - like choosing a movie, buying a car, selecting a stock portfolio, etc. - the choice should not be random or biased by someone's suggestion. MCDM algorithms often produce conflicting results when compared together because of the choice of the function to optimize.

This comes from the unavoidable trade-off be-

tween conflicting objectives as well as constraints. As a result, the optimal solution is not unique and corresponds to a so-called Pareto solution (Freund and Williamson, 2015).

On the opposite, learning to rank (LTR) is a class of approaches that apply supervised machine learning (ML) to resolve ranking problems. The training data for a LTR model consists of a list of samples and a "ground truth" score for each of those samples, manually labeled by experts, see (Li et al., 2017). The set of ranked data ("ground truth") becomes the data set that the system "trains" by minimizing some loss function to learn how best to rank automatically these items (Chaudhuri and Tewari, 2015).

The most common application of LTR is search engine ranking (Sharma et al., 2022). We propose to use them in the context of sports medicine because performance evaluation is a kind of fuzzy task.

Existing LTR algorithms may be divided into 3 main classes: (a) pointwise methods which reduce the rating on each item to regression or classification (Blackburn and Ukhov, 2013) (b) pairwise methods which essentially formulate ranking on each document pair as a classification problem (Borges et al., 2007) (c) list-wise methods which optimize a measure-specific loss function, on all available items. See Chavhan *et al.* (Chavhan et al., 2021) for a review.

The pros and cons of using LTR vs. MCDM are (a) LTR is essentially a black box in terms of explainability. It's hard to explain what exact effect specific inputs have on the outcome (b) LTR is

<sup>a</sup> <https://orcid.org/0000-0001-5917-6041>

<sup>b</sup> <https://orcid.org/0000-0002-1192-7333>

\*This research was supported by the program *Cátedras Franco-Brasileiras no Estado de São Paulo*, an initiative of the French consulate and the state of São Paulo (Brazil). We thank our colleagues Rémi Souriau for his helpful comments and Laurence Le-Moyec who supervise the data acquisition with the Institut national du sport, de l'expertise et de la performance (INSEP).

greedy ( $c$ ) result relevance is metric-dependent. This work presents a new extension of the LTR based on a continuous restricted Boltzmann machine (CRBM) (Hinton, 2002). The CRBM is a generative stochastic artificial neural networks (ANN) that can learn a probability distribution over (all possible permutations of) its set of inputs. CRBMs have found applications in dimensionality reduction (Vrábel et al., 2020), classification (Yin et al., 2018) and collaborative filtering (Verma et al., 2019). They are excellent generative learning models for latent space extraction. Specifically, they can be trained into excellent ranking devices because of their flexible loss function and the associative memory captured in the transfer matrix  $W$  between the visible and the hidden layers, which is a promising advantage over other standard MCDM algorithms.

The paper is organized as follows: section 2 provides a detailed description of the problem and the metrics used for measuring aggregation of ranks. Section 3 presents the generative model. The experimental results are presented in Section 4. The last section concludes and outlines the way for future work.

**Notations.** Throughout this paper small Latin letters  $a, b, \dots$  represent integers. Small bold letters  $\mathbf{a}, \mathbf{b}$  are put for vectors, and capital letters  $A, B$  for matrices or tensors depending on the context. The dot product between two vectors is denoted  $\langle \mathbf{a}, \mathbf{b} \rangle$ . We denote by  $\|\mathbf{a}\| = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$ , the  $\ell_2$  norm of a vector.  $X_1, \dots, X_n$  are non ordered variates,  $x_1, \dots, x_n$  non ordered observations. "Ordered statistics" means either  $p_{(1)} \leq \dots \leq p_{(n)}$  (ordered variates) or  $p_{(1)} \leq \dots \leq p_{(n)}$  (ordered observations). The  $p_{(i)}$  are necessarily *dependent* because of the inequality relations among them.

**Definition 1** ((Savage, 1956)). *The rank order corresponding to the  $n$  distinct numbers  $x_1, \dots, x_n$  is the vector  $t = (t_1, \dots, t_n)^T$  where  $t_i$  is the number of  $x_j$ 's  $\leq x_i$  and  $i \neq j$ .*

The rank order  $t$  is always unambiguously defined as a permutation of the first  $n$  integers.

## 2 GENERAL FRAMEWORK

### 2.1 Rank-Aggregation

Let  $A = \{a_1, a_2, \dots, a_n\}$  be a set of alternatives, candidates, individuals, etc. with cardinality  $|A| = n$  and let  $V$  be a set of voters, judges, criteria, etc. with  $|V| = m$ . The data is collected in a  $(n \times m)$  table  $T$  of general term  $\{t_{ij}\}$  crossing the sets  $A$  and  $V$  (Figure 1).  $t_{ij}$

can be marks ( $t_{ij} \in \mathbb{N}$ ), value scales ( $t_{ij} \in \mathbb{R}$ ), ranks (such that a voter can give *ex-aequo* positions) or binary numbers ( $t_{ij} \in \{0, 1\}$  such as opinion yes/no).  $T$  represents the ranking of the  $n$  alternatives under the form (see (Brüggenmann and Patil, 2011) for a reminder on rank-aggregation). For ease of writing, in the following,  $t_{ij} = t_i^{(j)}$ .

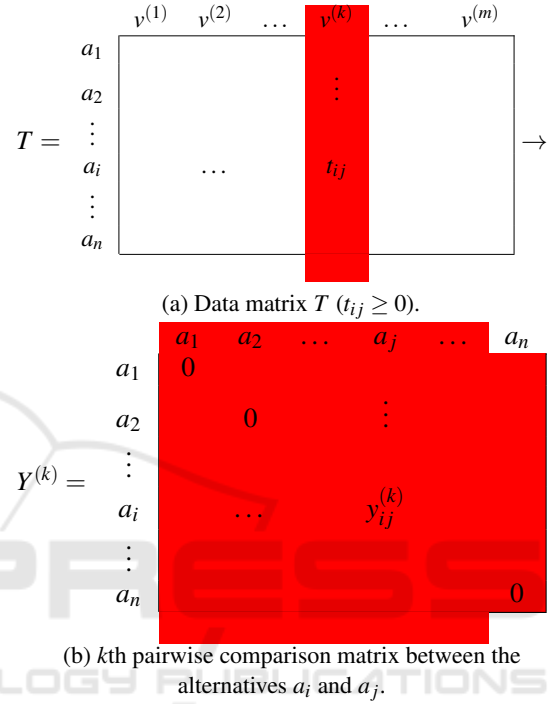


Figure 1: The data are collected in a  $(n \times m)$  table  $T$ .

Solving a rank-aggregation problem means finding a distribution of values  $x^*$  attributed by a virtual judge to the  $n$  alternatives by minimizing the disagreements of opinions between the  $m$  judges (Benson, 2016), i.e.

$$x^* = \arg \min_t \sum_{k=1}^m d(t, t^{(k)}), \quad \text{s.t. } t \geq 0, \quad (1)$$

where  $d(t, t^{(k)})$  is a metric measuring the proximity between  $t$  and  $t^{(k)}$ , chosen a priori, and  $t^{(k)}$  is the  $k$ th column of the table  $T$ . Depending on the properties of  $d(\cdot)$ , we will deal with a nonlinear optimization program with an explicit or implicit solution.

One could also stand the dual problem of the previous one, i.e., is there a distribution of rankings/marks that the  $m$  voters could have attributed to a virtual alternative 'a' summarizing the behavior of the set of individuals  $A$  (Yadav and Kumar, 2015)? The first problem is linked to the idea of aggregating points of view, and the second to the concept of summarizing behaviors.

## 2.2 Explicit or Implicit Resolution

Eq. (1) defines a nonlinear optimization program whose solution is  $x^*$  (Yadav and Kumar, 2015). The distance  $d(t^{(k)}, t^{(k')})$  between the ranking of voters  $k$  and  $k'$  can be chosen for instance as the Euclidean distance  $\sum_{i=1}^n (t_{ik} - t_{ik'})^2$ , the *disagreement* distance (Condorcet)  $\sum_{i=1}^n \text{sgn} |t_{ik} - t_{ik'}|$  or the order disagreement distance  $\sum_i \sum_j |y_{ij}^{(k)} - y_{ij}^{(k')}|$  as  $t^{(k)}$  can be replaced by its permutation matrix  $Y^{(k)}$  (Figure 1). In the latter,  $y_{ij}^{(k)} = \mathbb{1}_{i < j}$  denotes the indicator matrix for which  $y_{ij}^{(k)} = 1$  if the rank of the alternative  $a_i$  is less than the alternative  $a_j$  and 0 otherwise (Gehrlein and Lepelley, 2011). Note that  $y_{ii}^{(k)} = 0$  and  $y_{ij}^{(k)} = 0$  if  $i$  and  $j$  are ex-aequos.

In using matrix  $Y^{(k)}$ ,  $\frac{1}{2} \sum_i \sum_j |y_{ij}^{(k)} - y_{ij}^{(k')}| = \frac{1}{2} \sum_i \sum_j (y_{ik}^{(k)} - y_{ik}^{(k')})^2$  since since the expressions  $|y_{ij}^{(k)} - y_{ij}^{(k')}|$  are 0 or 1.

As  $y_{ij}^2 = y_{ij} = y_{ij}^{(k)^2} = y_{ij}^{(k)} = 0$  or 1, the function associated to order disagreement distance  $d$  is given by

$$\frac{1}{2} \left[ \sum_{i=1}^n \sum_{j=1}^n m y_{ij} + \sum_{i=1}^n \sum_{j=1}^n \left( \sum_{k=1}^m y_{ij} \right) - 2 \sum_{i=1}^n \sum_{j=1}^n y_{ij} \sum_{k=1}^m y_{ij}^{(k)} \right]. \quad (2)$$

Let  $\alpha_{ij} = \sum_{k=1}^m y_{ij}^{(k)}$  the total number of voters preferring alternative  $a_i$  to  $a_j$  and define a matrix  $A = \{\alpha_{ij}\}$ , summing the  $m$  matrices  $Y^{(k)}$  associated to the rankings  $t^{(k)}$  of the voters  $V^{(k)}$ , Eq. (2) becomes:

$$\frac{1}{2} \left[ \sum_{i=1}^n \sum_{j=1}^n m y_{ij} + \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} - 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} y_{ij} \right]. \quad (3)$$

Finally, the search for a total order given by a matrix  $Y$  is the solution of the linear program

$$\begin{aligned} \max_Y & \left( \sum_{i=1}^n \sum_{j=1}^n m y_{ij} + \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} - 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} y_{ij} \right) \\ \text{s.t.} \quad & \alpha_{ij} = \sum_{k=1}^p y_{ij}^{(k)} y_{ij} + y_{ji} = 1, i < j, \\ & y_{ii} = 0 \quad y_{ij} + y_{ji} - y_{ik} \leq 1, i \neq j \neq k, y_{ij} \in \{0, 1\}. \end{aligned} \quad (4)$$

If the chi-2 metric is chosen, then the dependent variables cannot be separated as in Eq. (4): The resolution follows an implicit gradient-descent procedure as in (Vigneron and Tomazeli Duarte, 2018). Section 2.3 detailed how chi-2 distance is used for solving aggregation problems.

## 2.3 Chi-2 Metric

The distance between  $x$  and  $t^{(k)}$  is given by

$$d(x, t^{(k)}) = \sum_i \frac{1}{f_i} \left[ \frac{f_{ix}}{f_x} - \frac{f_{ik}}{f_k} \right]^2, \quad (5)$$

where

$$\begin{aligned} f_i &= \frac{\sum_k t_{ik} + x_i}{\sum_{ik} t_{ik} + \sum_i x_i} \\ f_{ik} &= \frac{x_i}{\sum_{ik} t_{ik} + \sum_i x_i} \\ f_x &= \frac{\sum_i x_i}{\sum_{ik} t_{ik} + \sum_i x_i} \\ f_{ik} &= \frac{t_{ik}}{\sum_{ik} t_{ik} + \sum_i x_i} \\ f_k &= \frac{\sum_i t_{ik}}{\sum_{ik} t_{ik} + \sum_i x_i}. \end{aligned} \quad (6)$$

Let  $n\bar{r} = \sum_{ik} t_{ik}$ ,  $n\bar{x} = \sum_i x_i$ ,  $t_{.k} = \sum_i t_{ik}$  and  $t_{i.} = \sum_k t_{ik}$ . Then, after some calculus, the optimal ranking  $x^*$  minimizes  $\sum_k d(x, t^{(k)})$ :

$$n \sum_{k=1}^m \sum_{i=1}^n \frac{\bar{r} + \bar{x}}{t_{i.} + x_i} \left[ \frac{x_i}{n\bar{x}} - \frac{t_{ik}}{t_{.k}} \right]^2 = n(\bar{r} + \bar{x}) \sum_{i=1}^n \sum_{k=1}^m \left[ \frac{x_i}{\sqrt{t_{i.} n\bar{x}}} - \frac{t_{ik}}{\sqrt{t_{i.} t_{.k}}} \right]^2, \quad (7)$$

assuming  $t_{i.} \gg x_i$  and  $\sqrt{t_{i.} + x_i} \approx \sqrt{t_{i.}}$ .

According to Eq. (7), the ranking is performed on the row profiles or column-profiles of the matrix  $T$  (see Fig. 1), each row being weighted by  $\sqrt{t_{i.}}$ .

So it is equivalent to compute profile matrix  $C$  whose entry is  $\frac{t_{ik}}{\sqrt{t_{i.} t_{.k}}}$ , to consider the Euclidean distance between its rows and to heighten each row by  $\sqrt{t_{i.}}$ . A remark has to be made at this stage: two alternatives will be close if a large proportion of judges choose them simultaneously. For example, if there is a considerable amount of individuals chosen preferably by judges 'A' and 'B,' then we will say that judges  $A$  and  $B$  are close and that they "attract" each other. Eq. (7) is the well-known expression used in testing for independence in contingency tables.

Eq. (7) derives from the Bhattacharyya directed divergence between two discrete probability distributions  $P = \{p_i\}$  and  $Q = \{q_i\}$  defined as  $BD = -\ln(\sum_i \sqrt{p_i q_i})$  (Nielsen, 2022) if  $p_i = \frac{x_i^2}{t_{i.} n^2 \bar{x}^2}$  and  $q_i = \frac{t_{i.}^2}{t_{i.} t_{.k}}$ . Note that  $n^2$  is useless in the ratio and will be removed in the entries of the continuous restricted Boltzmann machine. Implicit methods are natural for LTR algorithms that are usually fed by an incoming data stream,  $n$  constantly varying. Section 3 proposes a learning model in which the rank probabilities take the form of a Boltzmann distribution.

### 3 METHODOLOGY

#### 3.1 Continuous Restricted Boltzmann Machine

Chen and Murray proposed another Boltzmann machine (BM) approach with *continuous* neuron in (Chen and Murray, 2003): the CRBM, a restricted Boltzmann machine using the neuron structure depicted in figure 2. In the CRBM, the activation function is unique for each neuron and given by:

$$s_j = \phi_j(X_j) = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + \exp(-a_j X_j)} \quad (8)$$

where  $\theta_L$  and  $\theta_H$  are, respectively, the function's lower and upper bounds.  $a_j$  is a slope parameter of  $\phi_j(\cdot)$ . The continuous behavior for the hidden units allows us to capture more information than binary units.

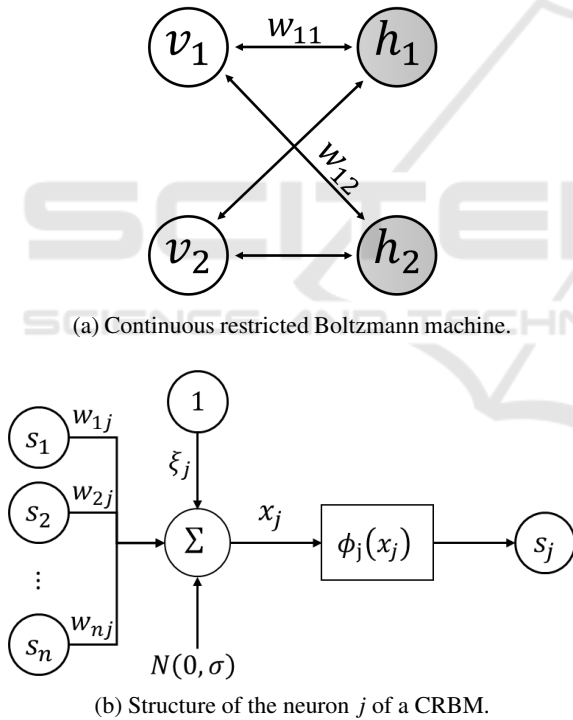


Figure 2: White neurons are visible neurons and gray neurons are hidden neurons in the CRBM. the coefficient  $w_{ij}$  refers to the weight of the symmetric link between the  $i$ -th visible unit  $v_i$  and the  $j$ -th hidden unit  $h_j$ .

We note  $W \in \mathbb{R}^{(m \times \ell)}$  the *transfer* matrix between the two layers and  $\xi^v$  and  $\xi^h$  the bias vectors of, respectively, the visible layer and the hidden layer. The energy function of the CRBM is:

$$E(v, h) = -v^T W h - v^T \xi^v - h^T \xi^h + \sum_i \frac{1}{a_i} \int_0^{s_i} \phi^{-1}(s') ds', \quad (9)$$

with  $\phi^{-1}(\cdot)$  the inverse of the activation for a coefficient slope  $a_i = 1$ . The energy  $E(s)$  of a CRBM is associated with the joint probability of the state of the neurons  $P_{\text{CRBM}}(s)$  defined as

$$P_{\text{CRBM}}(s) = \frac{1}{Z} \exp(-E(s)), \quad (10)$$

where  $Z$  is a marginalization constant. Training a CRBM is performed in minimizing the energy function Eq. (9), which itself requires sampling the hidden units.

The CRBM training uses the contrastive divergence algorithm (see (Hinton, 2012)). The training set  $\mathcal{D} = \{v^k\}_{1 \leq k \leq n}$  is composed of  $n$  observations used to find the best set of parameters  $\mathcal{P} = \{W, \xi\}$ ,  $\xi$  regrouping visible and hidden bias vectors.

Minimizing directly the joint log-likelihood  $\sum_{k=1}^n \log P_{\text{CRBM}}(v^k)$  to update the parameters is difficult due to the presence of the constant  $Z$ . Then it is replaced by the minimization of the contrastive divergence (MCD) (Hinton, 2002) that minimizes the contrast  $D$  between two successive Kullback-Leibler (KL)-divergences:

$$D = KL(P^0(v), P^\infty(v)) - KL(P^q(v), P^\infty(v)), \quad (11)$$

where  $P^0(v)$ ,  $P^\infty(v)$ ,  $P^q(v)$  are the distribution function of the visible units over respectively the training set, the equilibrium state and after  $q$  steps of Gibbs sampling (Hinton, 2012) (Fig. 3).



Figure 3: An intuitive idea is to minimize the KL divergence between  $P^0(v)$  and  $P^\infty(v)$ . But  $P^\infty(v)$  is intractable. We prefer to minimize  $D$ . If  $D = 0$ , then  $P^0(v) = P^1(v)$  and then :  $P^0(v) = P^\infty(v)$ .

An important observation is that any linear combination of measures of discrepancy with positive coefficients is also a measure of discrepancy.

$$KL(P^0(v), P^\infty(v)) - KL(P^q(v), P^\infty(v)) + \lambda(BD^0(v) - BD^q(v)), \quad (12)$$

with  $\lambda$  a regularization parameter. And thus, Eq. (12) can be used as a measure of discrepancy.

In particular, the observations are normalized:  $v = (\frac{t_{i,1}^2}{t_{i,1}^2}, \dots, \frac{t_{i,m}^2}{t_{i,m}^2})^T$  (see section 2.3).

In the next section, a CRBM driven by the loss function (12) ranks rugby players according to their performances measured by metabolomics.

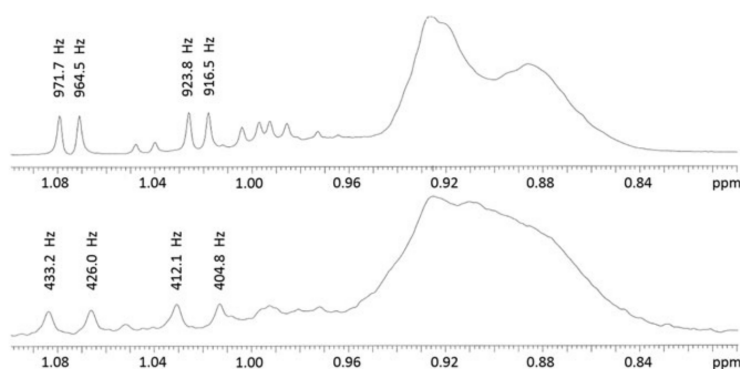


Figure 4:  $^1\text{H}$  NMR spectrum of human plasma acquired at 900 (top) and 400 MHz (bottom), from (Louis et al., 2017).

## 4 EXPERIMENTS WITH METABOLOMICS DATA

Endurance is a widely practiced sporting activity, from novice to champion. It is defined as maintaining an effort for a prolonged period. This effort originates from significant physiological and metabolic stress leading to organism adaptations. If this effort is too great, it can cause metabolic and locomotor disorders. The objective is to optimize training methods that will protect the health of athletes, young or old, efficient or less efficient. We adopt an integrative approach that simultaneously studies the physiological responses to exercise and the molecular and metabolic signals. From multivariate statistical analysis of biofluids such as urine, serum, plasma, saliva, sweat, etc., it is possible to generate metabolomic profiles or biomarkers. See (Khoramipour et al., 2022) for a review of metabolomics practice in sports medicine.

Since the 00's metabolomics investigates quantitatively the metabolome of living systems in response to pathophysiological stimuli or genetic modification (Amara et al., 2022).

Nuclear magnetic resonance (NMR) is traditionally used to elucidate molecular structures. It takes advantage of the energy transition of nuclear spins in a strong magnetic field to identify and elucidate the structure of organic molecules and specific metabolites. Metabolites are intermediate organic compounds resulting from metabolism. To understand the metabolomic changes induced by endurance exercise and training of rugby players according to the intensity and duration of the activity, we study the physiological modulation of rugby players according to their positions.

The study focuses on the activity variability between the forwards - more intense and intermittent efforts - and the rears - greater distance covered, more running, more rest time, etc. (Paul et al., 2022). It

aims to answer whether, during matches, (a) the urinary metabolites are identical before and after 80 minutes of a match? (b) this metabolomic modulation is of the same order depending on the player's position?

No study to date has investigated how to predict physiological exertion in rugby or how to classify a player according to its physiological parameters.

The experimental protocol is as follows: the urine of 80 players (40 forwards and 40 rears) is analyzed by NMR to identify the metabolites present in the two situations described above. The variations in metabolism explain the variations in physiological parameters as a function of the time and position factors. NMR spectra contain more than 10,000 values. See Fig 4 for an example of NMR spectrum.

19 variables represent the physiological variables, among which: forward/backward position of the player during the match, body mass index, experience, playing time, distance covered on the playground but also plasmatic metabolite rates in phenylalanine, tyrosine, glucose, creatinine,  $\beta$ -hydroxybutyrate, lactate, pyruvate, *N*-acetyl glycoprotein, lipids (Table 1). This set of variables constitutes the criteria that are used to rank the rugby players.

There are 70 samples in the training set and 10 in the test set. The structure of the CRBM is: 19 visible and 3 hidden units. Due to the small training and test data, we did not divide the data into mini-batches during the experiment. All the data were divided into eight groups for the seven-fold cross-validation method. Seven groups were selected as the training set each time, and the remaining group was the test set. This process was repeated until each group became a test set. The number of iterations was 200 for each CRBM. For training, the CRBMs were initialized with small random weights and zero bias parameters. The learning rate was  $\eta = 0.1$  when training with CD in Eq. (12) and  $\lambda = 0.005$ . CRBMs models

Table 1: Data description. Notice that variable 17 precises the position of the players. Forwards: pillars, hooker, 2nd lines. Backward: 3rd lines, 9 and 10, backs.

	Variables	type (if real: $\mu \pm \text{std.err.}$ )	physiological signification
1	term	binary	1st term=1, 2nd term=0
2	P	binary	position forward=1, backward=0
3	A	$28,4 \pm 0.97$	Age
4	H	$179.95 \pm 1.0$	height
5	W	$89.75 \pm 2.81$	weight
6	BMI	$27.65 \pm 0.756$	body-mass index
7	X	$41,3 \pm 1.01$	game experience
8	Ty	$143,9 \pm 1.89$	tyrosine
9	Glu	$269,5 \pm 2.59$	glucose
10	$\beta$	$414,1 \pm 3.22$	$\beta$ -hydroxybutyrate
11	Creat	$6.57 \pm 0.22$	creatinine
12	Gly	$5069 \pm 11.25$	glycoprotein
13	D	$2.238 \pm 7.48$	distance covered on the playground
14	L	$51,6 \pm 1.13$	lipids
16	R	$\{1,2,3\}$	rolling position
17	PHC0	$-86.98 \pm 15.73$	phenylalanine 0
18	PHC1	$23.36 \pm 9.19$	phenylalanine 1
19	SNR	$41.03 \pm 3.56$	signal to noise ratio

Table 2: Test results comparing CRBM with classic MCDMs algorithms VIKOR and ELECTRE IV. Float numbers are issued from the models and the ranks in sorting these numbers.

Players	VIKOR		ELECTRE IV		CRBM+Bhattacharyya	
1	3.608	<b>7</b>	3.127	<b>1</b>	3.291	<b>1</b>
2	5.123	<b>1</b>	6.324	<b>6</b>	6.074	<b>7</b>
3	4.639	<b>6</b>	5.094	<b>8</b>	4.925	<b>8</b>
4	5.378	<b>8</b>	4.923	<b>7</b>	5.088	<b>6</b>
5	5.811	<b>3</b>	6.147	<b>4</b>	5.886	<b>3</b>
6	4.033	<b>2</b>	3.401	<b>3</b>	3.693	<b>4</b>
7	3.468	<b>4</b>	3.567	<b>10</b>	3.467	<b>5</b>
8	4.254	<b>10</b>	3.411	<b>5</b>	3.663	<b>10</b>
9	5.6220	<b>9</b>	6.2400	<b>9</b>	5.9600	<b>9</b>
10	5.8110	<b>5</b>	6.3240	<b>2</b>	6.0740	<b>2</b>

are updated while the new training example is completed.

The results obtained with VIKOR, ELECTRE IV, and the CRBM are gathered in Table 2 for comparison. For each MCDM algorithm, the first column measures the discrepancy of the model Eq. (12) and the second column the rank of the rugby player (in bold). They provide an interpretation that the CRBM method is closer to the actual results and far from the ELECTRE IV method (Roy, 1985). VIKOR is a MCDMs, ranking preferences among a set of alternatives in the presence of conflicting criteria under the concept of group regrets (Guiwu et al., 2020). ELECTRE IV assumes that all requirements (actually pseudo-criteria) have the same importance.

At first glance, the rankings are not similar but not so different either. The rankings order the players ac-

cording to the sportive qualities recorded in Tab. 1. For example, ELECTRE IV and CRBM would select player one first, then player 10. But ELECTRE IV would prefer player 6 in the third place while CRBM would choose player 5.

The discriminant representation provided by the hidden layer is the most determining factor in favor of CRBM against VIKOR or ELECTRE. The players were ranked by the first hidden neuron of the CRBM. The scatter-plot Figure 5 uses the values of the first two hidden neurons. This means the hidden neurons capture a latent representation capable of discriminating between the two classes, "rears" and "forwards."

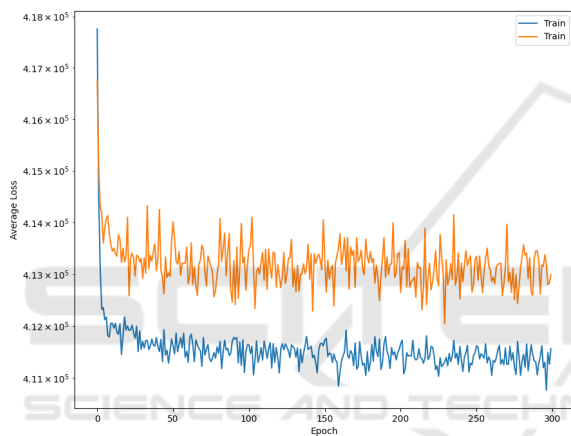
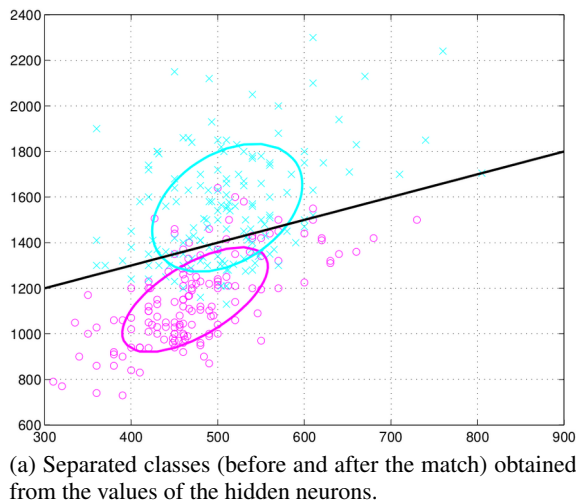


Figure 5: Score plot obtained with the two hidden neurons of the restricted Boltzmann machine (RBM): the samples separate each other before and after the match.

## 5 CONCLUSION AND DISCUSSIONS

CRBMs are domain-independent feature extractor that transforms raw data into latent variables. The most relevant questions are: how to dimension the hidden layer  $h$  optimally? And how do the neurons interact?

Our generative network is relatively small. Hence to compute  $p(h|v)$  is an affordable problem for small RBM, but once we have a large number of hidden neurons, it becomes impossible to compute all possible  $p(h|v)$ . The more neurons, the more computational efforts are needed: massive networks should not be the only way to reduce the modeling error. The

choice of dimension remains today an unsolved issue.

In addition, for each configuration  $v$ , some hidden neurons have a probability close to 0 or 1, meaning that for each  $v$ , some states of  $h$  are irrelevant.

Besides being energy-consuming, a significant dimension network requires much time to learn. In many papers, authors focus on comparing the performance between models but barely reach computational efforts between models. The issue of computational efforts can have a significant impact, particularly in the real-time system. Still, it is enormously dependent on the data, the application, and the used hardware.

To reduce the bias between the data distribution  $P_{data}(x)$  and the estimated data distribution  $P_{model}(x)$  the cost function was modified (Eq. 12).

The next step is constructing a deeper network, such as deep belief network (DBN), that may provide more explainability hints.

## REFERENCES

- Amara, A., Frainay, C., Jourdan, F., Naake, T., Neumann, S., Novoa-Del-Toro, E. M., Salek, R., Salzer, L., Scharfenberg, S., and Witting, M. (2022). Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Frontiers in Molecular Biosciences*, 9.
- Benson, D. (2016). *Representations of Elementary Abelian  $p$ -Groups and Vector Bundles*. Cambridge tracts in mathematics. Cambridge University Press, 1 edition.
- Blackburn, D. and Ukhov, A. (2013). Individual vs. aggregate preferences: The case of a small fish in a big pond. *Management Science*, 59:470–484.
- Brüggemann, R. and Patil, G. (2011). *Ranking and Prioritization for Multi-indicator Systems: Introduction to Partial Order Applications*. Environmental and Ecological Statistics. Springer-Verlag New York.
- Burges, C., Ragno, R., and Le, Q. (2007). Learning to rank with nonsmooth cost functions. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press.
- Chaudhuri, S. and Tewari, A. (2015). Perceptron like algorithms for online learning to rank. *ArXiv*, abs/1508.00842.
- Chavhan, S., Raghuwanshi, M. M., and Dharmik, R. C. (2021). Information retrieval using machine learning for ranking: A review. *Journal of Physics: Conference Series*, 1913(1):012150.
- Chen, H. and Murray, A. (2003). Continuous restricted boltzmann machine with an implementable training algorithm. *IEE Proceedings-Vision, Image and Signal Processing*, 150(3):153–158.
- Freund, D. and Williamson, D. (2015). Rank aggregation: New bounds for MCx. *CoRR*, abs/1510.00738.

- Gehrlein, W. and Lepelley, D. (2011). *Voting Paradoxes and Group Coherence: The Condorcet Efficiency of Voting Rules*. Studies in Choice and Welfare. Springer-Verlag Berlin Heidelberg, 1 edition.
- Guiwu, W., Jie, W., Jianping, L., Jiang, W., Cun, W., Fuad, E., and Tasawar, H. (2020). Vikor method for multiple criteria group decision making under 2-tuple linguistic neutrosophic environment. *Economic Research-Ekonomska Istraživanja*, 33(1):3185–3208.
- Hinton, G. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.
- Hinton, G. (2012). A practical guide to training restricted boltzmann machines. In Montavon, G., Orr, G. B., and Müller, K.-R., editors, *Neural Networks: Tricks of the Trade (2nd ed.)*, volume 7700, pages 599–619. Springer.
- Khoramipour, K., Sandbakk, O., Keshteli, A., Gaeini, A., Wishart, D., and Chamari, K. (2022). Metabolomics in exercise and sports: A systematic review. *Sports Med.*, 52(3):547–583.
- Li, X., Wang, X., and Xiao, G. (2017). A comparative study of rank aggregation methods for partial and top ranked lists in genomics applications. *Briefings in Bioinformatics*, 20(1):178–189.
- Louis, E., Cantrelle, F., Mesotten, L., Reekmans, G., and Adriaenssens, P. (2017). Metabolic phenotyping of human plasma by (1) h-nmr at high and medium magnetic field strengths: a case study for lung cancer. *Magnetic resonance in chemistry : MRC*, 55.
- Nielsen, F. (2022). Statistical divergences between densities of truncated exponential families with nested supports: Duo bregman and duo jensen divergences. *Entropy*, 24(3).
- Paul, L., Naughton, M., Jones, B., Davidow, D., Patel, A., Lambert, M., and Hendricks, S. (2022). Quantifying collision frequency and intensity in rugby union and rugby sevens: A systematic review. *Sports Medicine - Open*, 8.
- Rahman, A., Sungyoung, L., and Tae, C. (2017). Accurate multi-criteria decision making methodology for recommending machine learning algorithm. *Expert Systems with Applications*, 71:257–278.
- Roy, B. (1985). *Methodologie multicritere d'aide a la decision*. Economica, Paris,.
- Savage, R. (1956). Contributions to the theory of rank-order statistics – the trend case. *The Annals of Mathematical Statistics*, 27(3):590–615.
- Sharma, P., Yadav, D., Thakur, R. N., Reddy, K., and Praveen, M. (2022). Web page ranking using web mining techniques: A comprehensive survey. *Mob. Inf. Syst.*, 2022.
- Thakkar, J. (2021). *Multi-Criteria Decision Making*. Springer, first edition.
- Verma, S., Patel, P., and Majumdar, A. (2019). Collaborative Filtering with Label Consistent Restricted Boltzmann Machine. *arXiv e-prints*, page arXiv:1910.07724.
- Vigneron, V. and Tomazeli Duarte, L. (2018). Rank-order principal components. A separation algorithm for ordinal data exploration. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–6.
- Vrábel, J., Pořízka, P., and Kaiser, J. (2020). Restricted boltzmann machine method for dimensionality reduction of large spectroscopic data. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 167:105849.
- Yadav, N., author. abd Yadav, A. and Kumar, M. (2015). *An Introduction to Neural Network Methods for Differential Equations*. Springer, Gurgaon, Haryana, India.
- Yazdani, M., Fomba, S., and Zaraté, P. (2017). A Decision Support System for Multiple Criteria Decision Making Problems. In *17th International Conference on Group Decision and Negotiation (GDN 217)*, pages 67–75, Stuttgart, Germany.
- Yin, J., Lv, J., Sang, Y., and Guo, J. (2018). Classification model of restricted boltzmann machine based on reconstruction error. *Neural Computing and Applications*, 29:1–16.