Journal Title 1 (2016) 1–5
IOS Press

Optimizing Predictability of Rating Scales by Differential Evolution for the Use by Collective Intelligent Information and Database Systems

Waldemar W. Koczkodaj [a,*,**], T. Kakiashvili [b] Feng Li [c] Alicja Wolny-Dominiak [d] Jolanta Masiak [e]
[a] *Computer Science, Laurentian University, Sudbury, Ontario P3E 2C6, Canada, E-mail: wkoczkodaj@cs.laurentian.ca*
[b] *Sudbury Therapy, Sudbury, Ontario P3E 4M5, Canada, Insert a complete correspondence (mailing) address*
[c] ....
[d] *Department of Statistical and Mathematical Methods in Economics, University of Economics in Katowice, ul. 1 Maja 50, 40–287 Katowice, Poland, E-mail: alicja.wolny-dominiak@uekat.pl*
[e] *Independent Neurophysiological Unit, Department of Psychiatry, Medical University of Lublin, Lublin, Poland, E-mail: jolantamasiak@wp.pl*

Abstract. In this study, differential evolution (DE) optimization is proposed for the rating scale predictability improvement. An arbitrary assignment of equal values for rating scale items is used as the classifier although domain experts are aware that the contribution of individual items may vary. Most academic examinations are conducted by the use of rating scales. Rating scales are also used in psychiatry (especially for screening). This study demonstrates that the differential evolution is effective for optimizing the predictability of rating scales.

Keywords: Rating scale, DE, keyword three, keyword four, keyword five

1. Introduction

Rating scales are designed for knowledge aquisition in order to rate an entity, objects, or concept. Rating scales are also called assessment scales. In our study, we use ”the scale” when no ambiguity occurs. Probably, the most significant and frequently used rating scales are examinations or tests (e.g., Ontario Driver's test with 40 multiple-choice questions). Evidently, the ramifications of giving a driving license to unqualified driver are crown example why collective intelligence is of considerable importance since the test is developed by a group of experts to accommodate social needs of nearly the entire population of the age 16 years and higher.

Many popular rating scales use values “1 to 10” but five "stars" are gaining popularity for the online reviewing of goods or services. In rating scales, yes/no answers are also used for answers. The number of questions (called items in the rating scale terminology) may drastically differ from scale to scale. Numerous rating scales have over 100 items to rate but one item rating scale is also useful for rating the customer's satisfaction with goods or services. An example of a popular rating scale is the intelligence quotient (IQ). It is a total score derived from several standardized tests designed to assess the intelligence of an individual per-

*Corresponding author. E-mail: wkoczkodaj@cs.laurentian.ca.
**Euro Grant: Human Capital.

Fig. 1. Rating scale example.

son. The collective IQ is often used interchangeably with the term collective intelligence.

Misnomers of the scale is are a *survey* and a *questionnaire*. A questionnaire is a tool for data gathering and may not be used for rating. A survey may not necessarily be conducted by questionnaires and usually does not rate anything. Its goal is to gather data without rating them. Some surveys may be conducted by interviews or extracted by Internet agents with or without our consent or knowledge. The important distinction of rating scales from questionnaires and surveys, is that the rating scales are used for assessments. It means that rating scales are expected to have an outcome making them classifiers (in the terminology of statistics and machine learning). The scale term in the rating scale has the meaning as in “the scale of disaster” hence this study assumes that:

$$[ratingscale] = [dataframe] + [assessment]$$

$$dataframe] + [assessment] = [classifier]$$

The ’assessment’ procedure must be in place for a questionnaire or survey to become a rating scale. The assessment procedure may be as complex as the imagination of their authors but most psychiatric rating scales use a simple summation. The simple summation has been used in the first examples. The simple summation can be replaced by, for example, the assignment of weights. In our example 2, we computed weights by the differential evolution (using R packagbe DEval). As expected, it has not influenced the reduction, but it has improved the predictability rate.

’A picture is worth a thousand words’ therefore Fig. 1 has been used to illustrate ratings scales which are used in many ratings of various products.

The importance of subjectivity processing was driven by the idea of *bounded rationality*, proposed by Herbert A. Simon (a Nobel Prize winner), as an alternative basis for the purely mathematical modeling of decision making. Objective data are more commonly used in strict sciences. The deficiency of methods for processing subjectivity is the main reason why subjectivity is avoided whenever it is possible. However, objectivity is illusive and there is a fine line between subjectivity and objectivity in practice. For example, an item listed for sale for, say 1,000 monetary units, will be very likely sold for 999 units if such offer is put forward to us. If so, one may also not resist 998 units and so on. Setting a limit (so called, “the bottom line”) is often a highly subjective decision.

To illustrate our goal, let us focus our attention on a University exam with several parts or problems. Each part is marked from 0 to 3. There is a good reason for using small numbers. It is addressed in [7] for pairwise comparisons. From the psychological point of view, a smaller marks give smaller losses of imperfect answers. For example, 2.5 of 3 is 0.5 marking point loss, Evidently, it is 2 when 12 points are used for marking but 2 of 12 "looks more than 0.5 (that is, "only" half a point) of 3. Regardless of the maximum mark used for marking a single problem, we use different multipliers for more difficult problems than for easier problems. The multipliers are optimized for passing the maximum students. Evidently, no optimization is needed to fail all students.

Rating scales are of considerable importance in psychiatry (see [8, 13]) where the use of medical test (such as blood or urine) is not helpful in determination of a mental disorder. Common examples of scales are the Likert scale and 1-10 rating scales in which a person selects a number reflecting the quality of an entity (e.g., pain) as he/she perceives it. It is not uncommon for a rating scale to have over 100 items (questions) to rate. The primary use of rating scales is screening but they may be also used for decision making (e.g., a preliminary diagnosis). We stress that rating scales involves processing of subjective data. It is the absence of a well-established unit (e.g., one kilogram or meter) that compels us to use rating scales. A common mistake is regarding rating scale as a questionnaire since not all questionnaires are used for ratings. Statistical surveys are examples of using questionnaires without expectation of any rating.Rating scales are of considerable importance in psychiatry (see [8, 13]) where the use of medical test (known as biomarkers discussed in [24, 25] is not helpful in determination of a mental disorder.

2. Area Under Curve (AUC) of Receiver Operator Characteristic (ROC) as the objective function

A standard optimization model is maximizing (or minimizing) one real-valued function f in the parameter-space $\vec{x} \in \mathcal{P}$, or its specified subset $\vec{x} \in D$ where D denotes the set defined by the constraints. The maximization of a real-valued function $g(x)$ is equivalent to the minimization of the function $f(x) := -g(x)$.

In our nonlinear optimization problems, the objective function f has a large number of local minima and maxima. In fact, there are sectors (ranges) of values for one variable with identical values as demonstrated by Fig. 2. Finding the global maximum of a function is far more difficult and evolutionary methods are often used.

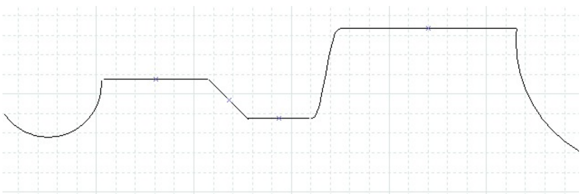

Fig. 2. Function with sectors of flatness

In case of a scale, we are optimizing *Area Under Curve of Receiver Operator Characteristic*. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The method was introduced during the World War II for evaluating the performance of radar operators. Since 1990s, the use of ROC exploded exponentially reaching approximately 15,000 citations by Web of Knowledge in 2016 and may be even higher in 2017. The main attraction of ROC is the accommodation of false positives and false negatives by a binary classifier. One of the best introductions to ROC curve constructions is [5] and readers unfamiliar with ROC are encouraged to get familiar with this source.

Different variations of meta-heuristic methods are popular for global optimization. In our case, we have many variables. The problem is that “flatness” may occur on many variables hence there is really no way of knowing which to go and every real interval consist of continuum of values. In case of a section, at the extrema, there are $\aleph_0$ solutions. Moreover, verifying the “mountain peaks” is impossible for n independent input variables since there are $\aleph_0$ candidate values, even if we set one input variable, and none of these values can be excluded from computations.

The most common is two-class prediction problem (binary classification). In medicine, it comes to "sick" or "not sick"; "dead" or "alive". In other words, a diagnosis is a mapping of observations into two classes. The diagnosis result can be a real (continuous output) or integer value, in which case the classifier boundary between classes must be determined by a threshold value (for instance, to determine whether a person has major depression) based on a total of a rating scale.

The diagnosis outcomes are labeled either as positive (P) or negative (N). There are four possible outcomes from a binary classifier. If the outcome from a prediction is P and the actual value is also P, then it is called a true positive (TP). However, if the actual value is N and the predicted value is P, then it is said to be a false positive (FP). Similarly, a true negative (TN) takes place when both the prediction outcome and the actual value are N. The false negative (FN) is when the prediction outcome is N while the actual value is P.

Consider a diagnostic test that seeks to determine whether a patient has a major depression. A false positive occurs when the patient’s total score of a scale indicates depression but actually the patient does not have the depression. A false negative, on the other hand, occurs when the patient’s total indicates lack of depression but the patient has the depression.

Let us define an experiment from P positive instances and N negative instances for some condition. The four outcomes can be formulated in a 2 by 2 *contingency table* or *confusion matrix* (CM), as follows:

$$\mathrm{CM} = \begin{bmatrix} \mathrm{TP} & \mathrm{FP} \\ \mathrm{TN} & \mathrm{FN} \end{bmatrix}$$

Sensitivity or *true positive rate* (TPR) is defined as

$$\mathrm{TPR} = \mathrm{TP}/\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}).$$

The *specificity* or *True Negative Rate* (TNR) is defined as

$$\mathrm{TNR} = \mathrm{TN}/\mathrm{N} = \mathrm{TN}/(\mathrm{FP} - \mathrm{TN}).$$

As stipulate earlier, [5] has compiled all basic terms and the above definitions. ROC curve plots the true positives (sensitivity) vs. false positives (1 specificity), for a binary classifier as its discrimination threshold is varied. A receiver operating characteristic (ROC) is used and evaluate the diagnostic (prognostic) performance of rating scales. The area under the curve

(AUC) of a receiver operating characteristic (ROC) curve reduces ROC performance to a single value representing the expected performance. AUC is equal to the probability that a classifier will rank a randomly chosen positive observation higher than a randomly chosen negative observation. It measures the classifier ability for ranking a set of observations according to the degree to which they belong to the positive class, but without actually assigning patterns to classes.

The term "receiver operating characteristic" was used in testing the ability of World War II radar operators to determine whether a blip on the radar screen represented an object (signal) or noise. This method, originally used by the signal detection theory, was later applied to psychiatry in [6]). The use of AUC as measure for the qualifier quality is addressed in [22].

Let us denote the probability for belonging in the positive class as a function of a decision/threshold parameter C_o as $p_1(C_o)$ and the probability of not belonging to the class as $p_0(C_o)$. The false positive rate FPR is given by:

$$\mathrm{FPR}(C) = \int_{p_0(C_o)}^{\infty} p_0(C)\, dC.$$

The true positive rate is defined as:

$$\mathrm{TPR}(C) = \int_{p_1(C_o)}^{\infty} p_1(C)\, dC.$$

The ROC curve plots parameters $\mathrm{TPR}(C)$ versus $\mathrm{FPR}(C)$ with C as the varying parameter, see e.g. [5].

2.1. Example

The goal of this example is to present the rating scale as the classifier defined as the weighted sum of attributes with weights $w_i = 1, 1 = 1, 2, \ldots 6,$. In that case the formula for this classifier is given by:

$$S_i = w_1 V_{i1} + ... + w_6 V_{i6} = V_{i1} + ... + V_{i6}, i = 1, ..., 26. \tag{1}$$

In that case it is possible to plot ROC curve and use the AUC measure of performance across all possible classification thresholds. We take in advantage an artificial data set placed in Tab. 1.

We have a class D with values 0 or 1 in the first column followed by six independent variables V_i, $1 = 1, 2, \ldots 6$, taking integer values, and their total sum S in the last column. Evidently, observations 7, 8, and 9 do not belong to class D since they have too high sum. Similarly, observations from 10 to 17 seem to be in the wrong class as they have the low total score.

Table 1

Sample data

#	D	V_1	V_2	V_3	V_4	V_5	V_6	S
1	0	0	0	0	0	0	1	1
2	0	0	1	0	0	0	1	2
3	0	0	1	2	0	0	0	3
4	0	0	0	0	0	0	3	3
5	0	0	1	1	0	1	0	3
6	0	0	2	1	0	0	0	3
7	0	2	2	0	3	2	2	11
8	0	1	3	1	2	2	3	12
9	0	2	3	3	3	0	3	14
10	1	0	0	1	0	1	1	3
11	1	1	0	0	0	1	1	3
12	1	0	0	1	0	0	2	3
13	1	0	1	1	0	1	1	4
14	1	0	2	2	0	0	0	4
15	1	0	1	1	0	1	1	4
16	1	1	2	1	0	0	0	4
17	1	2	2	3	3	1	2	13
18	1	2	0	2	3	3	3	13
19	1	2	3	3	2	2	2	14
20	1	2	2	2	2	3	3	14
21	1	2	2	3	2	2	3	14
22	1	3	3	2	3	1	3	15
23	1	2	3	3	3	1	3	15
24	1	3	3	3	2	2	3	16
25	1	3	2	3	2	3	3	16
26	1	3	3	2	3	3	3	17

In order to plot the ROC curve, we tread the column D as the known target, the independent variables V_i, as attributes and the column S as the classifier. The result is shown in Fig. 3.

The AUC for classifier S is equal $AUC = 0.8007$, which means a good accuracy of this classification. In case of a rating scale, we can say that the predictability of this scale is good enough. In the following, we propose a method of optimization of weights for the S classifier using the DE algorithm, what is illustrated in Example 2.

3. A brute force search for better weights

For computing AUC, we use two and only two columns (vectors) from Table 1. They are D (decision)

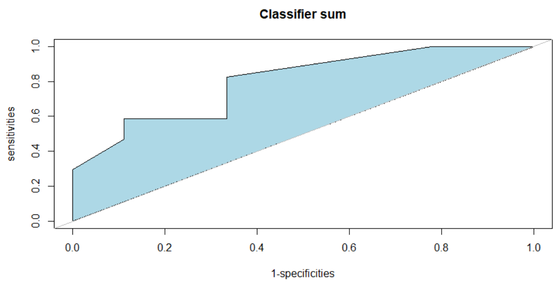

Fig. 3. ROC curve - sum (weights equal to 1)

and S (sum). Adding all independent variables V_i to get $S = \sum_{i=1}^{6} V_i$ makes a silent and often unreasonable assumption that all variables V_i are of equal contribution to diagnosis (expressed by all weights w are equal to 1). We know that in practice, it is not true and computing values w by maximizing AUC is needed. The optimizing problem may be formulate as follows.

Maximize AUC for a vector D and $S = \sum_{i=1}^{6} w_i V_i$ by changing w_i:

1. Run 10,000,000 the brute force Monte Carlo.
2. Localize "flatness" for each variable (it may have 0, 1 or more).
3. Find the "optimum point" on each flatness.
4. Run the brute force Monte Carlo for plus/minus 10% from the above point (20%?)

It is worth noting that the weights w_i are set for the entire data. In our case, a vector of weights was randomly generated for plus/minus 50% of the given initial values of 1 and AUC computed for them.

Practical experimentation revealed that the force brute approach is not acceptable. Even for a moderate number of variables (say 7), there are 7^{10} combinations taking already minutes on a fast Intel computer to compute AUC for each random vector w. Evidently, is not statistically acceptable to check only 10 random values for each variable.

4. A better way of optimizing weights by Differential Evolution

Differential evolution (DE), introduced in [23] and recently ued in [19] finds a solution by iterative improvement of a candidate solution (e.g., eights) with regard to a given measure of quality (e.g., AUC value). It is one of the most powerful meta-heuristics algorithms that operates on the basis of the same developmental process in evolutionary algorithms (EAs). Nevertheless, different from traditional EAs, DE uses the scaled differences of vectors to produce new candidate solutions in the population. Hence, no separate probability distribution should be used to perturb the population members [3]. The DE is also characterized by the advantages of having few parameters and ease of implementation. The application of DE on engineering [1] and biomedical [2] studies has attracted a high level of interest, concerning its potential.

Basically, the DE algorithm works through a particular sequence of stages. First, it create an initial population that sampled uniformly at random within the search bounds. Thereafter, three components namely mutation, crossover and selection are adopted to evolve the initial population. The mutation and crossover are used to create new solutions, while selection determines the solutions that will breed a new generation. The algorithm remains inside a loop until stopping criteria are met. In the following, we explain each stage separately in a more detail.

4.1. Initialization

Like other optimization algorithms, DE starts with a randomly initialized population of order NP consisting of parameter vectors the so-called individuals. Each such individual represents a D-dimensional vector of decision variables. The *ith* individual of the population at generation G could be denoted as follows:

$$\overrightarrow{X}_{G,i} = [x_{G_{i,j}}] = [x^G_{i,1}, \ldots, x^G_{i,D}], \tag{2}$$

where $j = 1 \ldots, D$ and $i = 1, \ldots, \mathrm{NP}$.

For each individual of the population, both upper and lower bounds of the decision variables should be restricted to their minimum and maximum values

$$\min \overrightarrow{X}_{G,i} = [\min x^G_{ij}] = [\min x^G_{i,1}, \ldots, \min x^G_{i,D}],$$

$$\max \overrightarrow{X}_{G,i} = [\max x^G_{ij}] = [\max x^G_{i,1}, \ldots, \max x^G_{i,D}]$$

Once initialization search ranges have been determined, DE assigns (at G=0) each individual a value from within the specified range as follows [23], for $G = 0$,

$$x^0_{i,j} = \min x^0_{i,j} - r(\max \overrightarrow{X}_{0,i} - \min \overrightarrow{X}_{0,i}), \tag{3}$$

where $r \in [0, 1]$ represents a uniformly distributed random number and NP the population size.

4.2. Mutation

After initialization, mutation operator produces new solutions by forming a mutant vector (trial vector) with respect to each parent individual (target vector). For each target vector, its corresponding trial vector can be generated by different mutation strategies. Each strategy employs different approaches to make a balance between the exploration and exploitation tendencies. For *ith* target vector at the G generation the five most well-known mutation strategies are presented as follows [18]. Here $r1, r2, r3, r4, r5 \in$ NP are five different randomly generated integer numbers. Furthermore, F is a scaling factor $\in [0, 2]$ affecting the difference vector and best $\in$ NP is an index of the best individual vector at generation G.

(1) DE/rand/1

$$\overrightarrow{V}_{G,i} = \overrightarrow{X}_{G,r1} + F(\overrightarrow{X}_{G,r2} - \overrightarrow{X}_{G,r3}), \tag{4}$$

(2) DE/best/1

$$\overrightarrow{V}_{G,i} = \overrightarrow{X}_{G,\text{best}} + F(\overrightarrow{X}_{G,r1} - \overrightarrow{X}_{G,r2}), \tag{5}$$

(3) DE/rand-to-best/1

$$\overrightarrow{V}_{G,i} = \overrightarrow{X}_{G,i} + F(\overrightarrow{X}_{G,\text{best}} - \overrightarrow{X}_{G,i}) +$$

$$+F(\overrightarrow{X}_{G,r1} - \overrightarrow{X}_{G,r2}), \tag{6}$$

(4) DE/best/2

$$\overrightarrow{V}_{G,i} = \overrightarrow{X}_{G,\text{best}} + F(\overrightarrow{X}_{G,r1} - \overrightarrow{X}_{G,r2}) +$$

$$+F(\overrightarrow{X}_{G,r3} - \overrightarrow{X}_{G,r4}), \tag{7}$$

(5) DE/rand/2

$$\overrightarrow{V}_{G,i} = \overrightarrow{X}_{G,r1} + F(\overrightarrow{X}_{G,r2} - \overrightarrow{X}_{G,r3}) +$$

$$+F(\overrightarrow{X}_{G,r4} - \overrightarrow{X}_{G,r5}). \tag{8}$$

4.3. Crossover

In this step, DE applies a discrete crossover approach to each pair of the parent vector and its corresponding trial vector. The basic version of DE incorporates the binomial crossover defined as follows [23]:

$$U_{G,i,j} = \begin{cases} V_{G,i,j} \text{ if } (\text{rand}_j[0,1) \leqslant CR) \text{ or } (j = j_{\text{rand}}) \\ X_{G,i,j} \text{ otherwise,} \end{cases}$$

where, CR is the user-specified crossover rate which determines the probability of mixing between parent and trial vectors. Also, $\text{rand}_j \in [0, D]$ is a randomly picked integer number.

4.4. Selection

In this step, DE adopts a selection mechanism to choose the best individuals according to their fitness for producing the next generation of population. Toward this goal, it compares performance of the trial and target vectors and copies the better one into next generation; as presented above.

$$\overrightarrow{U}_{G+1,i} = \begin{cases} \overrightarrow{U}_{G,i} \text{ if } f(\overrightarrow{U}_{G,i}) \leqslant f(\overrightarrow{X}_{G,i}) \\ \overrightarrow{X}_{G,i} \text{ otherwise,} \end{cases} \tag{9}$$

where, f is the objective function that should be optimized. To sum up, a detailed pseudo code of the aforementioned DE steps is presented in Figure 4.

Sliders are used for data collection by the Internet. Domain experts know the range for each variables and may be guided by results of the data analysis. The growing/decreasing tendency is what usually is needed to make a decision if a variable could be set to a certain value. In the past, integers were used for weights only because it was easier to do the calculations. There is no longer such reason and real values can be collected by a slider instead of a classic paper version of the Likert scale [16], The Internet version of it is the use of a slider represented by Fig. 1.

4.5. Example

This example illustrated optimizing the predictability of rating scales shown in Example 2.1 by optimizing the weights of the classifier S 1. The objective function is taken as the $AUC(w_1, ..., w_6)$ measure of performance. In determine the optimum of the objec-

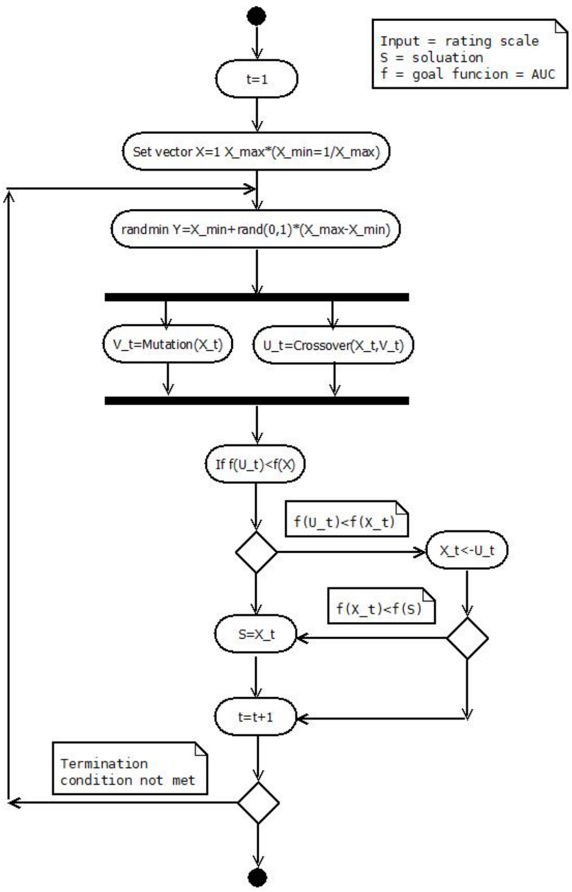

Fig. 4. Pseudo-code of the DE algorithm

tive function we use the DE algorithm shown in Fig 4 with the number of iteration as 200. The optimization are carried out with the package *optimDE* of the R program. Table 2 presents the results of the optimization, it is *maxAUC* and corresponding weight.

Table 2

Results of DE optimization for artificial data

	weight.opt
w_1^{DE}	0.91
w_2^{DE}	1.33
w_3^{DE}	2.75
w_4^{DE}	2.42
w_5^{DE}	2.93
w_6^{DE}	2.72
max AUC	0.863

In Fig. 5 we plot following graphs (from left hand side):

- ROC curve for classifier S with weights w_i^{DE}, $1 = 1, 2, \ldots 6$, - the rating scale predictability measures by AUC is 0.863
- ROC curve for classifier S with weights $w_i = 1$, $1 = 1, 2, \ldots 6$, - the rating scale predictability measures by AUC is 0.8
- comparison of two ROC curves

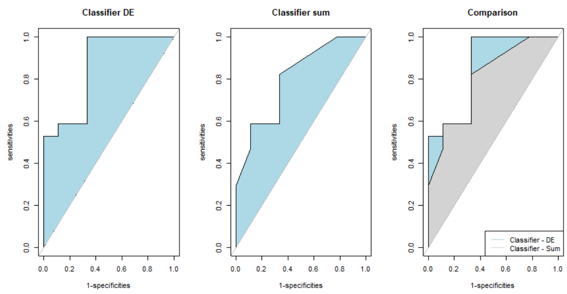

Fig. 5. ROC curves - comparison

The comparison plot clearly shows the higher area under ROC curve for classifier S with weights w_i^{DE}, $1 = 1, 2, \ldots 6$.

5. Dataset examples

An example has been used to demonstrate how to find the optimal weights for a given rating scale which was used for the Somerville Happiness Survey in 2015. The data are real (posted on the Internet *pro publico bono* [10]). The DE optimization is carried out using R package *DEoptim* (see [20]).

Table 3

Results of DE optimization for SHS data

	weight.opt
w_1^{DE}	1.15
w_2^{DE}	1.59
w_3^{DE}	1.14
w_4^{DE}	2.20
w_5^{DE}	0.42
w_6^{DE}	1.22
max AUC DE	0.704
AUC sum	0.679

Fig. 6 has been used to illustrate ROC curves for both classifier separately and comparison plot.

Differential evolution was used with 200 iterations. The weights with max AUC are in Table 3. The lower

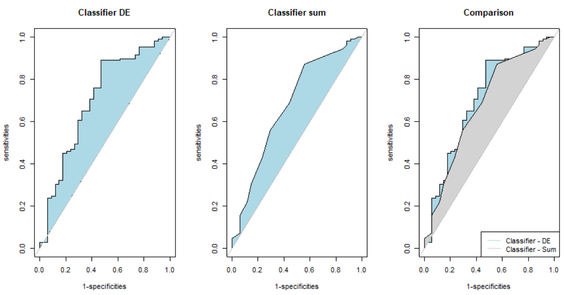

Fig. 6. ROC curves - comparison for SHS data

and upper bound decide are the search space of parameters. In our case, the lower bound is 0.1 and upper bound is 3.

6. Conclusions

The presented method is self-contained. However, it works well with [11] based on [12] and can be further improved by the pairwise comparisons (PCs). The proof o the PCs method convergence was provided in [9] and its statistical accuracy enhancement was analyzed in [14]. It has been implemented in R. However, it may be perceived a second step of the rating scale design process. The first step is the rating scale reduction described in [15] where a scale o 21 items was reduced to 6 items without the loss of predictability. The item reduction is not only an essential saving for (often expensive) data collection but also a contributor to the data collection error reduction. The more data are collected, the more errors are expected in them.

Acknowledgment

The first author has been supported in part by the Euro Research grant "Human Capital". Authors would like to thank Grant O. Duncan, Team Lead, Business Intelligence and Software Integration, Health Sciences North, Sudbury, Ontario) for his comments.

References

[1] Ankaiah, N.; Vadlamani, R. Multi objective constrained optimisation of data envelopment analysis by differential evolution. *International Journal of Information and Decision Sciences* 7(4): 334–357, 2015.

[2] Chowdhury, A.; Rakshit, P.; Konar, A. Protein-protein interaction network prediction using stochastic learning automata induced differential evolution. *Applied Soft Computing* 49:699-724, 2016.

[3] Das, S.; Suganthan, P. N. Differential evolution: a survey of the state-of-the-art. *IEEE transactions on evolutionary computation* 15(1): 4–31, 2011.

[4] Diagnostic and Statistical Manual of Mental Disorders, 5th Revision, American Psychiatric Association, 2013.

[5] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* 27(8):861–874, 2006.

[6] Ford,DE. Principles of Screening Applied to Psychiatric-Disorders. General Hospital Psychiatry Volume: 10 Issue: 3 Pages: 177-188,Published: MAY 1988.

[7] Fülöp, J.; Koczkodaj, W. W.; Szarek, S. J. A Different Perspective on a Scale for Pairwise Comparisons. N.T. Nguyen and R. Kowalczyk (Eds.): Transactions on CCI I, LNCS 6220: 71-84, 2010.

[8] Gan, Y.; Kakiashvili, T.; Koczkodaj, W. W.; Li, F. A note on relevance of diagnostic classification and rating scales used in psychiatry. *Computer Methods and Programs Biomedicine* 112(1):16–21, 2013.

[9] Holsztynski, W; Koczkodaj, WW Convergence of inconsistency algorithms for the pairwise comparisons Information Processing Letters, 59(4): 197-202, 1996.

[10] https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015

[11] Koczkodaj, Waldemar W.; Li, Feng; Wolny-Dominiak, Alicja, RatingScaleReduction package: stepwise rating scale item reduction without predictability loss R Journal, 10(1): 43-55: 2018.

[12] Koczkodaj, W.W. ; Kakiashvili, T. ; Szymanska, A. ; Montero-Marin, J. ; Araya, R. ; Garcia-Campayo, J.; Rutkowski, K.; Strzalka, D., How to reduce the number of rating scale items without predictability loss? Scientometrics, 111(2): 581-593, 2017.

[13] Kakiashvili, T.; Koczkodaj, W. W.; Woodbury-Smith, M. Improving the medical scale predictability by the pairwise comparisons method: Evidence from a clinical data study. *Computer Methods and Programs in Biomedicine* 105(3):210–216, 2012.

[14] Koczkodaj, WW Testing the accuracy enhancement of pairwise comparisons by a Monte Carlo experiment Journal of Statistical Planning and Inference, 69(1): 21-31, 1998.

[15] Koczkodaj, W. W.; Kakiashvili, T.; Szymańska, A.; Montero-Marin, J.; Araya, R.; Garcia-Campayo, J.; Rutkowski, K.; Strzałka, D. How to reduce the number of rating scale items without predictability loss? *Scientometrics* 111(2):581–593, 2017.

[16] Likert, R. A Technique for the measurement of attitudes. *Archives of Psychology* 140:1–55, 1932.

[17] Mental health action plan 2013 - 2020, World Health Organization, 2016.

[18] Qin, A. K.;, Huang, V. L.; Suganthan, P. N. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation* 13(2):398-417, 2009.

[19] Rakhshani, H.; Idoumghar, L.; Lepagnot, J.; Brévilliers, M. Speed up differential evolution for computationally expensive protein structure prediction problems. *Swarm and Evolutionary Computation*, 2019.

[20] Mullen, K.; Ardia, D.; Gil, D.; Windover, D.; Cline, J. DEoptim: An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software, Articles* 40(6):1–26, 2011.

[21] Rugg, H.O. Self-improvement of teachers through self-rating: A new scale for rating teachers' efficiency. *The Elementary School Journal* 20(9):670–684, 1920.

[22] Sonego, P.; Kocsor, A., Pongor., S. ROC analysis: applications to the classification of biological sequences and 3D structures. *Briefings in Bioinformatics* 9(3):198–209, 2008.

[23] Storn, R.; Price, K. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11(4):341–359, 1997.

[24] Wojcik, G. M., Masiak, J., Kawiak, A., Kwasniewicz, L., Schneider, P., Polak, N., & Gajos-Balinska, A., (2018). Mapping the human brain in frequency band analysis of brain cortex electroencephalographic activity for selected psychiatric disorders. Frontiers in neuroinformatics, 12.

[25] Wojcik, G. M., Masiak, J., Kawiak, A., Schneider, P., Kwasniewicz, L., Polak, N., & Gajos-Balinska, A. (2018). New Protocol for Quantitative Analysis of Brain Cortex Electroencephalographic Activity in Patients With Psychiatric Disorders. Frontiers in neuroinformatics, 12.