University of Glasgow

Chowdhury, S., Poet, R., and Mackenzie, L. (2013) *Multicriteria optimization to select images as passwords in recognition based graphical authentication systems.* In: HCI International 2013, 21 - 26 Jul 2013, Las Vegas, NV, USA.

http://eprints.gla.ac.uk/82561/

Deposited on: 31 March 2014

Multicriteria Optimization to Select Images as Passwords in Recognition Based Graphical Authentication Systems

Soumyadeb Chowdhury, Ron Poet and Lewis Mackenzie

School of Computing Science, University of Glasgow

{soumc, ron, lewis} @ dcs.gla.ac.uk

Abstract. Usability and guessability are two conflicting criteria in assessing the suitability of an image to be used as password in the recognition based graphical authentication systems (RGBSs). We present the first work in this area that uses a new approach, which effectively integrates a series of techniques in order to rank images taking into account the values obtained for each of the dimensions of usability and guessability, from two user studies. Our approach uses fuzzy numbers to deal with non commensurable criteria and compares two multicriteria optimization methods namely, TOPSIS and VIKOR. The results suggest that VIKOR method is the most applicable to make an objective statement about which image type is better suited to be used as password. The paper also discusses some improvements that could be done to improve the ranking assessment.

Keywords: image password, TOPSIS, VIKOR, multiple criteria analysis

1 Introduction

Recognition-based graphical authentication systems are an alternative type of mechanism where images are used as passwords. The decision making process to select the most suitable image type to be used as passwords in RBGSs has to consider several conflicting criteria (usability and guessability). There has been no existing research to quantify and rank the different image types suitable to be used as passwords in RBGSs. In this paper we propose a new approach that effectively integrates a series of techniques and concepts so that the decision makers can obtain a comprehensive and consistent evaluation result. Our approach combines: (1) multi criteria decision making (MCDM); (2) fuzzy set theory and fuzzy numbers; (3) a multi criteria optimization method.

MCDM [1, 2] usually helps decision makers quantify and evaluate each criterion as well as rank all the alternatives. The main steps of the decision making process involve []: (1) Establishing the evaluation criteria; (2) Consider the alternatives; (3) Assessing the alternatives in terms of criteria; (4) Employing multi criteria analysis technique; (5) Accepting one alternative as an optimal choice.

This research incorporates fuzzy set theory [3] in the performance measurement. The main contribution of fuzzy set theory is its ability to represent vague data. In a deci-

sion making process each criterion is measured with a different scale, which makes the judgments unbalanced and imprecise. Fuzzy numbers help to deal with the imprecise and uncertain values of each criterion [4]. Each fuzzy number is associated to a linguistic variable as shown in fig 1 [5].

Fuzzy number	Linguistic term
$\overline{1}$	Very poor
$\overline{3}$	Poor
$\overline{5}$	Ordinary
$\overline{7}$	Excellent
$\overline{9}$	Very excellent

Fig. 1. Fuzzy numbers with linguistic variables defining them

Multicriteria optimization is the process of determining the most feasible alternative of all taking in to account each of the established criteria. An approach to determine a final solution as a compromise was introduced in [6]. In this context, the VIKOR method [7, 8, 9] is a multicriteria ranking approach developed to help solve the decision problems with conflicting and noncommensurable criteria. It determines the compromise solution for a problem with conflicting criteria. A compromise solution is a feasible solution that is closest to the ideal which is agreed by mutual concessions. Another multicriteria ranking approach, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [10, 11] uses a distance based MCDM approach to choose an alternative that has the shortest geometric distance from the ideal solution and the longest geometric distance from the negative ideal solution. The ideal solution is the best performance value exhibited by any alternative for each criteria and negative ideal solution is the worst performance value.

The rest of the paper is structured as follows. In Section 2 we discuss the usability and guessability study conducted by us. In Section 3 we describe the ranking framework and compare the TOPSIS and VIKOR method with an illustrative example. The data for the example is obtained from the user studies discussed in Section 2. Finally, in Section 4 we discuss the results and scope for future improvements

2 User study

We developed four RBGSs (online website which used images as passwords). Each of them used a different image type as the password: (1) Mikon; (2) Doodle; (3) Art; (4) Object. We conducted a usability study with independent measures (between subjects) style of experimental design having four conditions namely Mikon, doodle, art and objects. The participants in the Mikon condition created four passwords (each password comprising of 4 Mikon images) and authenticated using them. The same procedure was followed in other conditions too. This study was conducted for eight weeks with 100 participants. The dependant variables were:

Memory: It examined the *average/mean successful login percentage (S)* for each of the conditions calculated as,

$$S = \frac{\text{Total number of successful login in the condition}}{\text{Total number of login in the condition}} \times 100$$

Time: It examined the average/mean registration time (Reg time), and average/mean login time of successful login (Log time). The average registration time for each condition was calculated as,

$$\frac{1}{4}\sum_{i=1}^{4} \text{Registration time for password } (i), \qquad \text{where } i \text{ denotes password } 1, 2, 3, 4$$

The average time of successful login for each condition was calculated as given below, z represents total number of successful login

$$\frac{1}{z}\sum_{n=1}^{z} \text{Login time for successful login } (n)$$

Exp-Use: This dimension had two parameters: (1) Sat-use i.e. the satisfaction scores; (2) Str-use i.e. the stress scores. Sat-use was assessed from the ratings (1- 5, 1 being highly dissatisfied to 5 being highly satisfied) given by the participants to the different aspects in the post study questionnaire- (sat1) Ease to register; (sat2) Ease to authenticate; (sat3) Meaningfulness/nameability of the image; (sat4) satisfaction with the type of image used as password. Similarly, str-use was assessed from the ratings (1- 5, 1 being least stressful to 5 being highly stressful) given by the participants: (str1) level of mental stress; (str2) level of physical stress; (str3) amount of effort required to choose images during registration; (str4) amount of effort required to successfully login.

A guessability study (70 participants) was conducted with the repeated measure protocol having 4 conditions for a period of 8 weeks: (1) login using Mikon passwords; (2) login using doodle passwords; (3) login using art passwords (4) login using object passwords. Each participant had to guess 4 passwords in each condition using the verbal descriptions of the image passwords given to them. Each condition was performed on a different day and participants were randomly assigned to the conditions. The parameters used to measure this criterion were:

Guessing: It examined the mean successful login percentage for each participant

Quality of the descriptions (descriptions): The participants were asked to rate some aspects on a scale of 1 to 5 (1 being low rating and 5 being high rating): q1-was the description easy to understand; q2-was the description adequate to identify the image; q3-was the description meaningful; q4-was the description useful.

Exp-Guess: This dimension had two parameters: (1) Sat-Guess i.e. the satisfaction scores; (2) Str-Guess i.e. the stress scores. For Sat-Guess, the participants were asked to rate on a scale of 1 to 5 (1 being lowest satisfaction and 5 being highest satisfaction): sat1- Ease to guess the passwords; sat2- Satisfaction of individual performance; sat3- Level of success in guessing the image; sat4- Efficiency to complete the task. In the case of Str-Guess, they were asked to rate on a scale of 1 to 5 (1 being lowest stress/effort and 5 being highest stress/effort): str1- Level of mental stress; str2- Level of physical stress; str3- Level of temporal stress; str4- Amount of effort required to guess.

3 Ranking Framework

Step1: Decision Matrix (Fig 2) - Calculate the mean value for each criterion in each alternative from the experiments. They are represented in the decision matrix as: mean value of the alternative A_i, where $i \leq n$ (n being the total number of alternatives) with respect to criteria C_j, where $j \leq m$ (m being the total number of criteria). Here the alternatives are the image types used as password.

Image types	Memory	Log time	Reg time	Sat-Use	Str-Use	Guessing	description	Sat-Guess	Str-Guess
Mikon	74.1	19.52	72.8	13.25	12.12	59.06	13.1	12.81	14.68
Doodle	67.04	22.16	75.4	12	13.12	74.28	14.6	14.84	12.35
Art	54.9	24.56	84.4	9.2	14.66	48.55	10.44	12.2	13.8
Object	77.3	18.28	70.6	13.91	11.87	87.32	16.44	15.9	8.27

Fig. 2. Mean scores for each dimension obtained from the experiments

Step 2: Judgement Matrix (Fig 5) - Each fuzzy number represents an interval for each criterion which is decided by the experimenter (Fig 3). Higher values are considered to be the best in the case of memory, sat-use, description and sat-guess, whereas lower values are considered to be the best in the case of guessing, reg time, log time, str-use and str-use. Then each in the decision matrix is replaced by the corresponding fuzzy number. Each fuzzy number is converted to a triangular fuzzy number T_{ij} using the membership function shown in Fig 4. Finally, the matrix is normalized using eq (1)

$$D_{ij} = \frac{T_{ij}}{\sqrt{\sum_{i=1}^{n} T_{ij}}} \qquad (1)$$

Fuzzy number	Memory	Reg time	Log time	Sat-Use	Str-Use	Guessing	description	Sat-Guess	Str-Guess
1	<45	≥ 90	>30	≤ 8	≥ 15	≥80	≥ 16	≥ 15	≤ 8
3	45-59	81-89	26-30	9-10	13-14	70-79	14-15	13-14	9-10
5	60-69	76-80	22-25	11-12	11-12	60-69	12-13	11-12	11-12
7	70-79	71-75	19-21	13-14	9-10	45-59	10-11	9-10	13-14
9	≥80	≤ 70	<19	≥ 15	≤ 8	<45	<10	≤ 8	≥ 15

Fig. 3. Interval values associated to fuzzy numbers in each criterion

Thus the judgement matrix (**D**) is obtained, which gives the overall judgement scores for each of the alternatives (A_i) with respect to criteria (C_j) without taking into account the relative weight between each criterion.

Fuzzy number	Membership Function
$\bar{1}$	(1,1,3)
$\bar{x}$	(x-2,x,x+2)
$\bar{9}$	(7,9,9)

Fig. 4. Membership function to convert singular to triangular fuzzy

	Memory	Time	Exp-Use	Guessing	description	Exp-Guess
Mikon	(0.64,0.61,0.58)	(0.55,0.54,0.55)	(0.66,0.61,0.58)	(0.69,0.67,0.64)	(0.16,0.32,0.39)	(0.66,0.64,0.6)
Doodle	(0.38,0.43,0.45)	(0.44,0.46,0.49)	(0.33,0.4,0.43)	(0.13,0.28,0.35)	(0.5,0.54.0.54)	(0.03,0.43,0.45)
Art	(0.13,0.26,0.32)	(0.22,0.31,0.37)	(0.41,0.3,0.36)	(0.69,0.67,0.64)	(0.83,0.76,0.7)	(0.66,0.64,0.6)
Object	(0.64,0.61,0.58)	(0.66,0.62,0.55)	(0.66,0.61,0.58)	(0.13,0.09,0.21)	(0.16,0.11,0.23)	(0.16,0.1,0.22)

Fig. 5. Judgement Matrix (D) with judgement scores for each image type

Step 3: Criteria Weights (Fig 6) - Determine the importance of each criterion by setting the criteria weights. This is done by the experimenter using subjective judgments. We assigned equal weights to usability and guessability (50% in each case).

Criteria	L	M	U
Memory	0.125	0.25	0.5
Time	0.075	0.15	0.3
Exp-Use	0.05	0.1	0.2
Guessing	0.125	0.25	0.5
Description	0.075	0.15	0.3
Exp-Guess	0.05	0.1	0.2

Fig. 5. Weights assigned to each criteria, L= (M/2) and U = (Mx2)

Step 4: Intermediate Performance Matrix (Fig. 7) - Each criterion weight is multiplied to the corresponding criterion score in the judgement matrix i.e. D_{ij} x W_j. The result is (F_{ij}), the performance score for each of the alternatives corresponding to the specified criteria.

Image type	Memory	Time	Exp-Use	Guessing	description	Exp-Guess
Mikon	(0.08,0.15,0.29)	(0.04,0.08,0.16)	(0.03,0.06,0.11)	(0.08,0.16.0.32)	(0.012,0.048,0.12)	(0.03,0.06,0.12)
Doodle	(0.04,0.1,0.22)	(0.03,0.07,0.14)	(0.01,0.01,0.08)	(0.01,0.07,0.17)	(0.037,0.08,0.16)	(0.001,0.04,0.09)
Art	(0.01,0.06,0.16)	(0.01,0.04,0.1)	(0.02,0.03,0.07)	(0.08,0.16,0.32)	(0.06,0.11,0.21)	(0.03,0.06,0.12)
Object	(0.08,0.15,0.29)	(0.05,0.09,0.16)	(0.03,0.06,0.11)	(0.01,0.02,0.1)	(0.01,0.016,0.07)	(0.008,0.01,0.044)

Fig. 7. Intermediate performance matrix with performance scores as triangular fuzzy numbers

Step 5: Defuzzification: First, the interval performance matrix (F^α) as in Fig 8 is obtained using the alpha cut method on the intermediate performance matrix (**F**). Each score is worked upon using eq. 2 and 3 to form an interval [F^α_{ijl} , F^α_{ijr}]. The value of α **(0-1)** denotes the experimenter's degree of confidence while evaluating the scores and criteria weights in the process. A larger value expresses stronger degree of confidence. Here, we set up the value to 0.85 because we are confident about the criteria weights chosen (equal distribution between usability and security).

$$F^\alpha_{ijl} = L_{ij} + \alpha \left(M_{ij} - L_{ij} \right) \qquad (2)$$

$$F^\alpha_{ijr} = U_{ij} + \alpha \left(U_{ij} - M_{ij} \right) \qquad (3)$$

	Memory	Time	Exp-Use	Guessing	description	Exp-Guess
Mikon	0.136	0.072	0.054	0.144	0.04	0.054
	0.178	0.096	0.07	0.192	0.06	0.072
Doodle	0.088	0.062	0.034	0.058	0.07	0.032
	0.124	0.084	0.048	0.09	0.09	0.05
Art	0.014	0.034	0.028	0.144	0.046	0.054
	0.08	0.052	0.038	0.192	0.13	0.072
Object	0.136	0.082	0.054	0.018	0.015	0.01
	0.178	0.104	0.07	0.036	0.021	0.016

Fig. 8. Interval performance matrix (F^α) obtained by applying $\alpha = 0.85$

The risk index β **(0-1)** represents the experimenter's positive/ negative view point about their experimental evaluation. A lower value represents positive optimism. Here we choose its value to be 0.15 because we are highly optimistic about our experimental evaluation, since it is reliable (can be reproduced) and valid (gives us significant results). The risk index is used together with the interval per-romance matrix to calculate the final performance matrix using eq. 4.

$$F^\alpha_{ij\beta} = \beta F^\alpha_{ijl} + (1 - \beta) F^\alpha_{ijr} \qquad (4)$$

	Memory	Time	Exp-Use	Guessing	description	Exp-Guess
Mikon	0.17	0.091	0.066	0.182	0.05	0.068
Doodle	0.115	0.079	0.045	0.083	0.08	0.046
Art	0.066	0.048	0.036	0.182	0.11	0.068
Object	0.17	0.099	0.066	0.032	0.02	0.014

Fig. 9. Final performance matrix ($\boldsymbol{F^{\alpha}_{\beta}}$) obtained by setting β =0.15

Step 5: Ideal and negative ideal solution (Fig. 10) - Determine the best value ($\boldsymbol{f^*}$) and the worst value ($\boldsymbol{f^-}$) for each criterion, using eq. 5 and 6 respectively.

$$f^* = \left\{ \max f_{ij} \right\} \quad (5) \qquad f^- = \left\{ \min f_{ij} \right\} \quad (6)$$

	Memory	Time	Exp-Use	Guessing	description	Exp-Guess
Ideal	0.17	0.099	0.066	0.182	0.11	0.068
Non Ideal	0.066	0.048	0.036	0.032	0.02	0.014

Fig. 10. Ideal and negative ideal solution for each criterion obtained from ($\boldsymbol{F^{\alpha}_{\beta}}$)

For VIKOR ranking: Compute the utility measure ($\boldsymbol{S_i}$) eq. 7, regret measure ($\boldsymbol{R_i}$) eq. 8 and the VIKOR index ($\boldsymbol{Q_i}$) eq. 9. Rank all the alternatives, sorting by the values utility measure, regret measure and Vikor index in decreasing order. The results are three ranking lists as shown in Fig 11.

$$S_i = \sum_{j=1}^{q} \frac{(f_j^* - f_{ij})}{(f_j^* - f_j^-)} \quad (7) \qquad\qquad R_i = Max \frac{(f_j^* - f_{ij})}{(f_j^* - f_j^-)} \quad (8)$$

$$Q_i = v \frac{S_i - S^*}{S^- - S^*} + (1 - v) \frac{R_i - R^*}{R^- - R^*} \quad (9)$$

$$S^* = Min\ (S_i),\ \ S^- = Max\ (S_i),\ \ R^* = Min\ (R_i),\ \ R^- = Max\ (R_i)\ \ and\ v\ is\ set\ to\ 0.5$$

The alternative with the minimum VIKOR index is best ranked if it satisfies the following two conditions:

Condition 1: Acceptable advantage

$$Q\ (rank:2) -\ Q\ (rank:1)\ \geq Z\ , where\ Z = \frac{1}{j - 1}$$

$$0.49 - 0.027 = 0.473 \geq\ \ 0.33\ \textbf{(Satisfied for Mikon)}$$

Condition 2: Acceptable stability in decision making- The alternative with rank 1 must also be best ranked by S or/and R. **(Satisfied for Mikon)**

	Utility measure	Rank list 1	Regret Measure	Rank list 2	Vikor Index	Rank list 3
Mikon	0.82	1	0.66	1	0.027	1
Doodle	2.98	2	0.7	2	0.49	2
Art	3	3	1	3	0.5	3
Object	3	3	1	3	0.5	3

Fig. 11. Three ranking lists produced by the VIKOR method

For TOPSIS ranking, compute the ideal separation S_i^+ (eq. 10), negative ideal separation S_i^- (eq. 11) and the relative closeness C_i (eq. 12). Then rank according to as shown in Fig 12.

$$S_i^+ = \sqrt{\sum_{j=1}^{m}(F_{ij} - F_{ij}^+)^2} \quad (10) \qquad S_i^- = \sqrt{\sum_{j=1}^{m}(F_{ij} - F_{ij}^-)^2} \quad (11)$$

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (12)$$

	Distance from ideal	Distance from negative ideal	Relative closeness to ideal	Rank list
Mikon	0.06	0.2	0.76	1
Doodle	0.12	0.10	0.45	3
Art	0.11	0.18	0.60	2
Object	0.18	0.11	0.39	4

Fig. 12. Ranking list for relative closeness to ideal in TOPSIS method

4   Discussion and Concluding remarks

The approach presented in the paper is the first work in the field of RBGSs (Human computer interaction-Security) to obtain a comprehensive and consistent evaluation result, while making a decision that considers two conflicting criteria, usability and guessability. The ranking given by VIKOR and TOPSIS are different. In case of TOPSIS, Mikon is ranked 1, followed by art, doodle, and object. This method considers the distance of an alternative from the ideal and negative ideal solution, without the relative importance of the distances, which is a major concern in the decision making process. Being far away from the negative ideal solution is not an essential advantage in decision making process and hence the importance of the separation

(distance) has to be considered. This is evident from Fig 9, where art has the best scores for guessability parameters, which is higher than the best scores of usability parameter for objects as well as doodles. Hence TOPSIS ranks art higher than doodles and objects. In case of VIKOR, the ranking is based on closeness to the ideal solution, which integrates the maximum group utility for the majority (eq. 7), and minimum individual regret for the opponent (eq. 8), together with acceptable advantage and stability. Here, Mikon is ranked 1 followed by doodle; objects (highest usability and highest guessability) are ranked the same as art (lowest guessability and lowest usability).

So the results suggest that a distance based ranking method may not always produce a solution which is closest to the ideal, especially when conflicting variables are being considered, which makes the ranking uncertain as well as unsuitable. But the VIKOR method has a very useful applicability in ranking the image types to be used as passwords in RBGSs, especially when a lot of conflicting variables are being considered for the decision making process. It is interesting to note that inclusion or exclusion of an alternative could affect the VIKOR ranking because the value of the VIKOR index depends on the ideal solution. In our ongoing work we are looking into various approaches to fix the ideal solution by defining some standards. In terms of improvement, the assessment of the criteria weights can be improved by integrating several expert's judgement and evaluate its degree of consistency.

References

1. Saaty, T.L. and Cho, Y (2001) The Decision by the US Congress on China's Trade Status: A Multicriteria Analysis. *Socio-Economic Planning Sciences*, vol. 35, pp. 243-252.

2. Lee, H. S. and Chou, M. (2006). T. A Fuzzy Multiple Criteria Decision Making Model for Airline competitiveness Evaluation. *Lecture notes in Computing Science*, No. 4252, pp. 902-909

3. Zimmermann, H. J. (2005). *Fuzzy Set Theory and its Applications*, fourth ed. Springer, New York.

4. Mikhailov, L. (2003) Deriving Priorities from Fuzzy Pairwise Comparison Judgments. *Fuzzy Sets and Systems*, vol. 134, pp. 365-385.

5. Saaty, T.L. (1980) *The Analytic Hierarchy Process*, New York: McGraw-Hill.

6. Zeleny, M. (1973). Compromise Programming, in Cochrane J.L. and M. Zeleny (Eds.), Multiple Criteria Decision Making, University of South Carolina Press, Columbia.

7. Opricovic, S. (1998). Multicriteria Optimization of Civil Engineering Systems. Technical report, Belgrade.

8. Tzeng, G., Teng, M., Chen, J. J. and Opricovic, S. (2002). Multicriteria Selection for a Restaurant Location in Taipei. *International Journal of Hospitality Management*. vol. 21 (2). pp. 171-187.

9. Tzeng, G., Lin, c. and Opricovic, S. (2005). Multicriteria Analysis of Alternative fuel Buses for Public Transportation. *Energy Policy*. vol. 33, pp. 1373-1383

10. Hwang, C. and Yoon, K. (1981). Multiple Attributes Decision making Methods and Application: a state-of-the-Art Survey. Springer, berlin Heidelberg.

11. Chu, T. (2002). Selecting Plant Location via a Fuzzy TOPSIS Approach. *International Journal of Advance Manufacturing and technology*. vol. 20, pp. 859-864.