A Method of User Preference Elicitation by Pairwise Comparisons

Alexander Borodinov
Geoinformatics and Information Security Department,
Samara National Research University,
Samara, Russia
aaborodinov@yandex.ru

Vladislav Myasnikov
Geoinformatics and Information Security Department,
Samara National Research University,
Samara, Russia
vmyas@geosamara.ru

***Abstract*—In this paper, we consider the problem of reconstructing functions defined implicitly by the results of pairwise comparisons. In the proposed approach, we apply an adaptive transformation to the high-dimensional space. Then we classify the comparisons using linear or non-linear classifiers. In this work, we consider linear regression and random forest as classification algorithms. In experimental analysis, we compare different methods of transformation to the high-dimensional space and investigate the effectiveness of the proposed method.**

***Keywords*—utility function, preference function, preferences elicitation, pairwise comparisons, machine learning**

I. Introduction

The method of pairwise comparisons is one of the methods used in recommendation systems. Analyzing pairwise comparisons, we try to determine some pattern in the choice of the preferred option. The method of pairwise comparisons uses information about comparing pairs of objects, in contrast to the classical methods of machine learning, which use data about a specific object [1-4]. The task of providing recommendations for a particular user is the task of preference elicitation.

Three main types of tasks are specified according to different types of objects and classes [3,5]:

- label ranking – search for preferred ordering among labels for any example. The traditional classification problem can be generalized as part of the label ranking problem when the classification result of the example is a label of the highest rank;

- instance ranking – ranking a set of examples for a fixed label order;

- object ranking – similar to ranking examples, however, labels are not associated with examples.

In this paper, we consider the task of ranking objects where the objects may be the transport routes proposed by the recommender system [6,7], and the preferences are the routes selected by the user. In the second section of the paper, we briefly describe the existing approaches to the construction of recommender systems. In the third section, we give the problem formulation and problem statement. The fourth section describes the method of pairwise comparisons. The fifth section shows the results of experimental studies. At the end of the work, conclusions and possible directions for further research are presented.

II. Related work

There is a large research community focused on recommendation systems with a wide range of tasks. Historically, most approaches are based on collaborative filtering approaches. For example, forecasting ratings for streaming services such as Netflix [8]. Pairwise methods for comparing user preferences are most often used in search engines [9]. Recommender systems based on information about transitions between sites and products in online stores are another large area of research [10,11]. One of the new approaches was the use of neural networks to improve the accuracy of recommendations [12]. One of the young and underdeveloped areas are transport recommender systems [13]. In our work, we consider the method of pairwise comparisons, which has not been used before for the construction of transport recommender systems.

III. Problem statement

Let the objects set $\Omega \equiv \{\omega_j\}_{j\in J}$ have an order «$\leq$» and/or a strict partial order «$\prec$». The equivalent notation is $\omega_i \succ \omega_j$ and $\omega_j \prec \omega_i$. In the case $\omega_i \leq \omega_j \wedge \omega_j \leq \omega_i$, the objects are indistinguishable and $\omega_i \ll \omega_j$. Absolute preference is characterized by utility function $u:\Omega \to R$, and relative preference is described by preference function $p:\Omega\times\Omega \to R$. For utility function $u(\omega_i) < u(\omega_j)$ denote as $\omega_i \prec \omega_j$, $u(\omega_i) \leq u(\omega_j) \Leftrightarrow \omega_i \leq \omega_j$ and $u(\omega_i) = u(\omega_j) \Leftrightarrow \omega_i \ll \omega_j$. For preference function $p(\omega_i,\omega_j) > 0$ denote as $\omega_i \succ \omega_j$ and $p(\omega_i,\omega_j) = 0 \Leftrightarrow \omega_i \ll \omega_j$. The preference function has restrictions based on the properties of the corresponding order relations such as asymmetry in argument, transitivity, etc.

A preference function can be defined through a utility function $p(\omega_j,\omega_i) = u(\omega_j) - u(\omega_i)$ and $u(\omega_j) = p(\omega_j,\omega^*)$ $\left(u(\omega^*) = p(\omega^*,\omega^*) = 0\right)$.

Objects are defined by the feature vector $\mathbf{x} \equiv \mathbf{x}(\omega) \in X$ of N-dimensional space. The utility and preference function will be written as $p(\mathbf{x},\mathbf{x}_j), u(\mathbf{x})$. Let $\mathbf{x}_j \equiv \mathbf{x}(\omega_j)$ and $p_{ij} \equiv p(\omega_i,\omega_j)$, $u_j \equiv u(\omega_j)$ to shorten the record. Information about pairwise comparisons can be presented in the form of values of the preference function $p(\omega_j, \omega_i)$ or in the form of a symbolic representation:

$$z_{ij} \equiv z(\omega_j, \omega_i) = \begin{cases} 1, & p(\omega_j, \omega_i) > 0, \\ 0, & p(\omega_j, \omega_i) = 0, \\ -1, & p(\omega_j, \omega_i) < 0. \end{cases}.$$

The choice of a specific route from the route list proposed by the system is an example of information on paired comparisons in a transport recommendation system.

The number of incorrectly reconstructed relationships, the Kendall distance for pairwise comparisons, is a criterion for the reconstruction quality of the preference and utility function:

$$d = \left| \left\{ (i,j) : z(\omega_i, \omega_j) \neq z(\mathbf{x}(\omega_i), \mathbf{x}(\omega_j)), \quad (i,j) \in I \right\} \right|,$$

This value in the normalized form is an estimate of the corresponding relation errors probability $\tilde{d} = d \cdot |I|^{-1}$.

IV. Method

A. Pairwise Comparison Method

Pair comparison methods were initially used to range objects that cannot be described by a feature vector. Each element of the matrix (c_{ij}) is the absolute frequency of the i-th object over j-th [14]. To analyze such data, the Thurstone model was proposed [15], in which it is assumed that the utility of an object is determined by a normally distributed random variable. Thus, for objects ω_0, ω_1 we get $f_u(u|\omega_j) \sim \mathrm{N}(\mu_j, \sigma_j^2)$, which with $u(\omega_1) - u(\omega_0) \sim \mathrm{N}(\mu_1 - \mu_0, \sigma_{10}^2), \quad \sigma_{10}^2 \equiv \sigma_1^2 + \sigma_0^2 - 2\rho_{10}\sigma_1\sigma_0$ and the Laplace function $\Phi(\cdots)$ we get:

$$P(\omega_1 \succ \omega_0) = P(u(\omega_1) - u(\omega_0) > 0) = \Phi\left(\frac{\mu_1 - \mu_0}{\sigma_{10}}\right).$$

In the numerical estimation of the probability (5) as the relative frequency of the corresponding preferences calculated using the matrix (c_{ij}), we have the following estimation:

$$\mu_1 - \mu_0 \triangleq \sigma_{10}\Phi^{-1}\left(\frac{c_{10}}{c_{10} + c_{01}}\right).$$

The simplified Thurstone model assumes the absence of correlation and equal variances in the utility function, which can be represented as: $\sigma_1^2 = \sigma_0^2 = 0.5, \quad \rho_{10} = 0, \quad \sigma_{10}^2 \equiv 1$.

Another featureless method is the Bradley-Terry model. [1]. Estimating the probability (5) in the following form:

$$P(\omega_1 \succ \omega_0) = \frac{\pi_1}{\pi_1 + \pi_0}, \quad \pi_j = \exp(\mu_j / s),$$

where s is a numerical non-negative parameter. Thus, we have the following estimate of the preferences between the objects:

$$\mu_1 - \mu_0 \triangleq s\left(\ln\left(\frac{c_{10}}{c_{10} + c_{01}}\right) - \ln\left(1 - \frac{c_{10}}{c_{10} + c_{01}}\right)\right).$$

The analytic hierarchy process (AHP) is used for the multicriteria ranking of objects that are defined by features. In this case, at the initial step matrices $(m_{ij}^n)_{i,j \in J}, \quad n = \overline{0, N-1}$ are calculated. Each element of the matrix is the result of a user response regarding the preferences of the i-th object over the j-th objects according to the n-th criterion. The resulting utility of the objects as a scalar product $u(\omega_j) = \sum_{n=0}^{N-1} w_n v_j^n$, where $\mathbf{v}^n = (v_0^n, \cdots, v_{|J|-1}^n)^T$ is the right eigenvector of the preference matrix, and w – eigenvector of the matrix of alternatives. The main problem of this method is a large number of pairwise comparisons. Therefore, in practice, they often use a model $u(\mathbf{x}(\omega)) = \sum_{n=0}^{N-1} w_n \cdot v_n(x_n(\omega))$, based on a generalized additive model.

B. Proposed method description

The following features should be considered when reconstructing the utility function and preference function:

- reconstruction of functions is practically impossible with a small amount of information or in its absence, as in the case of a system cold start;

- it must be able to automatically transform to nonlinear models. Classes will be separable almost surely when using transformation the original features space to a new feature space Y with a higher dimension;

- the regression task of reconstructing the utility function can be reduced to the classification problem by reconstructing the symbolic representation.

The method of function reconstruction by their symbolic representation contains the following steps:

– feature values normalization in the range [0,1];

– selection of a new feature space (basis) Y;

– transformation of the original feature vector $\mathbf{x}$ into the new feature space Y with a higher dimension K=dim(Y) ≥ N;

– building a linear or nonlinear classifier in the feature space Y;

– quality assessment of the building classifier on the test dataset.

In the case when the evaluation of the preference function is unsatisfactory, go to the selection of a new basis and transformation of the feature space.

The described steps are presented as a diagram in Fig. 1.

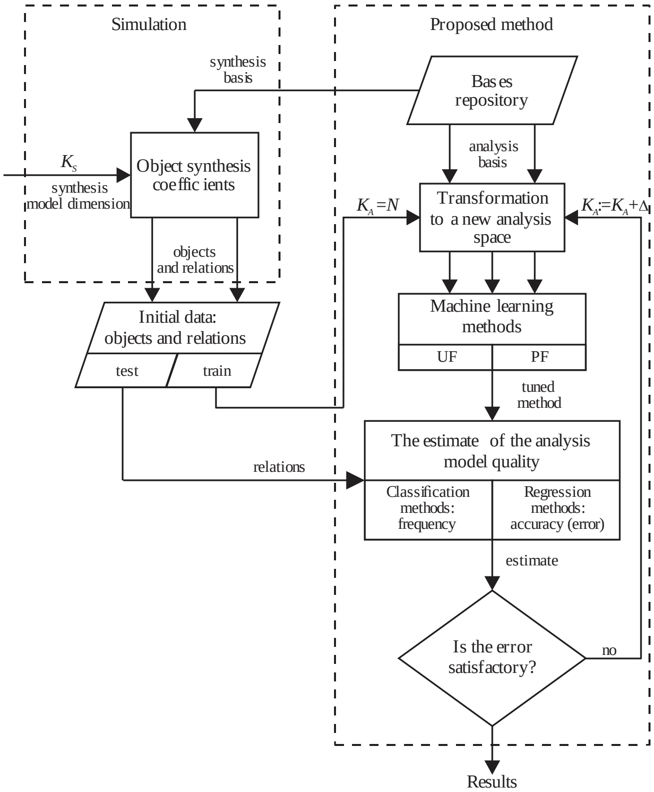

Fig. 1. The scheme of the proposed approach.

C. Bases Repository

In this paper, we consider the following bases for transformations $\varphi : \underset{\mathbf{x}}{X} \underset{\mapsto}{\rightarrow} \underset{\mathbf{y}(\mathbf{x})}{Y}$:

– Original basis:

$$K = \dim(Y) = \dim(X) = N, \quad y_n = x_n, \quad n = \overline{0, N-1};$$

– Polynomial basis:

$$K = \sum_{n=0}^{N-1} K_n = [K_n = K_0] = K_0 N = \dim(Y) > \dim(X) = N,$$

$$y_k = \prod_{n=0}^{N-1} x_n^{k_n}, \quad n = \overline{0, N-1}; \quad k = \sum_{n=0}^{N-1} K_0^n k_n;$$

– Fourier basis:

$$K = \sum_{n=0}^{N-1} K_n = [K_n = K_0] = K_0 N = \dim(Y) > \dim(X) = N,$$

$$y_k = \prod_{n=0}^{N-1} \cos(\pi k_n x_n), \quad n = \overline{0, N-1}; \quad k = \sum_{n=0}^{N-1} K_0^n k_n.$$

- Haar basis:

$$K = \sum_{n=0}^{N-1} K_n = [K_n = K_0] = K_0 N = \dim(Y) > \dim(X) = N,$$

$$y_k = \prod_{n=0}^{N-1} \varphi(x_n), \quad \varphi(x_n) = \sqrt{2^{k_n}} \varphi(2^j x - i),$$

$$i = 0, 1, \ldots, 2^j - 1, \quad j = 0, 1, \ldots, \log_2 N - 1,$$

$$n = \overline{0, N-1}; \quad k = \sum_{n=0}^{N-1} K_0^n k_n.$$

D. Machine Learning Methods

In this paper, we use logistic regression and random forest when testing the proposed approach.

Logistic regression solves the binary classification problem using a linear dividing hyperplane:

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_N .$$

Classifier parameters for a particular training set $\{\mathbf{x}_j, r_j\}_{j \in J}$ are determined from the condition:

$$J(\mathbf{w}) = \sum_{j \in J} \ln\left(1 + \exp\left(-r_j \cdot d(\mathbf{x}_j)\right)\right) \to \min ,$$

where $r_j \in \{-1, 1\}$ – is a random variable of the correct classification that determines the true class of the corresponding j-th object.

A random forest is a voting method implementation of several tree classifiers. A random forest avoids retraining, unlike a decision tree. Each tree is built independently of the rest on a random subset of the training set. The components of the feature vector are selected from a random subset of features for each partition when learning trees.

User decisions may be erroneous, especially with a small difference in the proposed alternatives. Therefore, in this work we use the Thurstone's model with the probability estimation to add errors in the ideal preferences. For the case $\mu_j > \mu_i$:

$$z_{ij} \Leftarrow \begin{cases} z(\omega_j, \omega_i), & rnd < P\left(u(\omega_j) - u(\omega_i) > 0\right), \\ -z(\omega_j, \omega_i), & otherwise. \end{cases}$$

where *rnd* <*R*[0,1] - random variable.

We train and test the model several times, averaging the results of the error calculation, in order to avoid the effect of unsuccessful partitioning of the set on the training and test datasets.

V. Experimental Research

We used the following parameters during the experiments:

– The synthesis model dimension Ks = 15, 35;

– The transformation model dimension Ka = 15, 35, 63;

– The paired comparisons number InstNum = 10000, 50000;

– The number of random partitions of the dataset nIter = 100.

Comparison of the effectiveness of various bases is presented in Table I.

TABLE I. THE COMPARISON RESULTS FOR FOURIER, POLYNOMIAL, AND HAAR BASES (LR –LOGISTIC REGRESSION, RF – RANDOM FOREST)

Ks	*Ka*	InstNum	Error					
			S: Fourier A: Polynomial		S: Fourier A: Haar		S: Polynomial A: Fourier	
			LR	RF	LR	RF	LR	RF
15	15	10000	0.2246	0.1129	0.0078	0.0482	0.0092	0.0115
15	15	50000	0.14208	0.06006	0.00172	0.02366	0.01004	0.00952
15	63	10000	0.1986	0.1159	0.0083	0.0509	0.008	0.0139
15	63	50000	0.15958	0.07988	0.00442	0.0319	0.00434	0.00792
Ks	*Ka*	InstNum	S: Polynomial A: Haar		S: Haar A: Fourier		S: Haar A: Polynomial	
			LR	RF	LR	RF	LR	RF
15	15	10000	0.0048	0.0101	0.0015	0.0408	0.0068	0.0067
15	15	50000	0.00164	0.0062	0.00222	0.0211	0.00244	0.0048
15	63	10000	0.0064	0.0091	0.0067	0.0422	0.0035	0.0051
15	63	50000	0.00144	0.00554	0.00248	0.02138	0.0014	0.00422

An increase in the number of pairwise comparisons led to a decrease in the error value, however, it significantly increased the program execution time, especially for the random forest method. We can state, based on the results, that the proposed approach has demonstrated efficiency and effectiveness.

VI. CONCLUSION

The paper proposes an approach to the reconstruction of functions defined implicitly by the results of pairwise comparisons. The approach is based on the transformation into the symbolic space of a greater dimension with the subsequent classification of the comparison results. It is shown that the proposed method allows us to effectively solve the problem of evaluating the user preference function. Logistic regression has a significant advantage in speed and stability. A further area of research is the application of the developed approach on data on preferred routes for users on public and private transport.

ACKNOWLEDGMENT

The work was funded by the Ministry of Science and Higher Education of the Russian Federation (unique project identifier RFMEFI57518X0177).

REFERENCES

[1] R.A. Bradley and M.E. Terry, "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," Biometrika, vol. 39, no. 3/4, pp. 324-345, 1952. DOI: 10.2307/2334029.

[2] P.C. Fishburn, "Utility theory for decision making," Huntington, N.Y: R.E. Krieger Pub. Co, 1979.

[3] J. Fürnkranz and E. Hüllermeier, "Preference Learning," Berlin Heidelberg: Springer-Verlag, 2011.

[4] K.P. Murphy, "Machine Learning: A Probabilistic Perspective," Cambridge, MA: The MIT Press, 2012.

[5] V.V. Myasnikov, "Reconstruction of functions and digital images using sign representations," Computer Optics, vol. 43, no. 6, pp. 1041-1052, 2019. DOI: 10.18287/2412-6179-2019-43-6-1041-1052.

[6] A.A. Agafonov, A.S. Yumaganov and V.V. Myasnikov, "Big data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a K nearest neighbors method," Computer Optics, vol. 42, no. 6, pp. 1101-1111, 2018. DOI: 10.18287/2412-6179-2018-42-6-1101-1111.

[7] A.A. Agafonov and V.V. Myasnikov, "Numerical route reservation method in the geoinformatic task of autonomous vehicle routing," Computer Optics, vol. 42, no. 5, pp. 912-920, 2018. DOI: 10.18287/2412-6179-2018-42-5-912-920.

[8] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," Computer, vol. 42, no. 8, pp. 30-37, 2009. DOI: 10.1109/MC.2009.263.

[9] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai and H. Li, "Learning to rank: From pairwise approach to listwise approach," ACM International Conference Proceeding Series, vol. 227, pp. 129-136, 2007. DOI: 10.1145/1273496.1273513.

[10] T. Joachims, "Optimizing search engines using clickthrough data," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 133-142, 2002.

[11] X. He, "Practical lessons from predicting clicks on ads at Facebook," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014. DOI: 10.1145/2648584.2648589.

[12] P. Covington, J. Adams and E. Sargin, "Deep neural networks for youtube recommendations," RecSys - Proceedings of the 10th ACM Conference on Recommender Systems, pp. 191-198, 2016. DOI: 10.1145/2959100.2959190.

[13] P. Campigotto, C. Rudloff, M. Leodolter and D. Bauer, "Personalized and Situation-Aware Multimodal Route Recommendations: The FAVOUR Algorithm," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 1, pp. 92-102, 2017. DOI: 10.1109/TITS.2016.2565643.

[14] K. Tsukida and M. Gupta, "How to Analyze Paired Comparison Data," 2011.

[15] L.L. Thurstone, "A law of comparative judgment," Psychological Review, vol. 34, no. 4, pp. 273-286, 1927. DOI: 10.1037/h0070288.