



Contents

Advancing Innovation and Development of IT and Quantitative Management: Preface for ITQM 2013 Y. Shi, Y. Xi, P. Wolcott, Y. Tian, J. Li, D. Berg, Z. Chen, E. Herrera-Viedma, G. Kou, H. Lee, Y. Peng, L. Yu	1
POS Data and your Demand Forecast J. Zhu	8
Relation based Access Control in Campus Social Network System Z. Du, Y. Liu, Y. Wang	14
China's Huji System Reform and Social Computing W. Wu	21
The Role of Text Pre-processing in Sentiment Analysis E. Haddi, X. Liu, Y. Shi	26
Improving Lap-TSVM with Successive Overrelaxation and Differential Evolution W.-J. Chen, Y.-H. Shao, Y.-F. Ye	33
Proximal Plane Clustering via Eigenvalues Y.-H. Shao, L. Bai, Z. Wang, X.-Y. Hua, N.-Y. Deng	41
Examining Disease Risk Factors by Mining Publicly Available Information J. Pedersen, F. Liu, F. Alfarraj, H. Ngondo	48
Metadata-based Information Resource Integration for Research Management Z. Chen, D. Wu, J. Lu, Y. Chen	54
Scene Image Mosaic based on Three Freedom Degree C. Liu, Y. Shi	62
Feature Selection with Attributes Clustering by Maximal Information Coefficient X. Zhao, W. Deng, Y. Shi	70
Pattern Discovery in DNS Query Traffic W. Ruan, Y. Liu, R. Zhao	80
A New Character Segmentation Approach for Off-Line Cursive Handwritten Words A. Choudhary, R. Rishi, S. Ahlawat	88
A Systematic Approach to Order Fulfillment of on-demand Delivery Service for Bento Industry R.-C. Chen, C.-H. Shieh, K.-T. Chan, S.-Y. Chiu, J.-Y. Fan, Y.-T. Chang, N.-J. Ma	96
A Unified Theoretical Framework for Data Mining D.M. Khan, N. Mohamudally, D.K.R. Babajee	104
Supervised Discretization with $GK - \tau$ W. Huang, Y. Pan, J. Wu	114
Document Clustering using Hybrid XOR Similarity Function for Efficient Software Component Reuse V. Radhakrishna, C. Srinivas, C.V. Guru Rao	121
Multi-criteria Risk Assessment of a Construction Project J. Tamošaitienė, E.K. Zavadskas, Z. Turskis	129
Towards the Definition of Relations between Clusters in Multicriteria Decision Aid Clustering B. Rouba, S.N. Bahloul	134
The Theory of Multi-criteria and Multiple Constraint-level Linear Programming for Oilfield Development Y. Zhong, D. Chen, G. Cui, Y. Jia, L. Li, Y. Liao	141
A Simple Decomposition Alternating Direction Method for Matrix Completion L. Niu, X. Zhao	149
Review of Multiple Criteria and Multiple Constraint-level Linear Programming D. Chen, Y. Zhong, Y. Liao, L. Li	158
A Data Mining Perspective of the Newsvendor Problem X. Yu, Z. Qi, Y. Zhao	166
Application of Multi-Objective Genetic Algorithm to Quotation of Global Garment Companies S.-S. Li, R.-C. Chen, Y.-H. Chen, M.-H. Wu, K.-H. Leng, H.-Y. Wang	173
Revised PROMETHEE II for Improving Efficiency in Emergency Response H. Zhao, Y. Peng, W. Li	181
User Acceptance Towards Web-based Learning Systems: Investigating the Role of Social, Organizational and Individual Factors in European Higher Education A. Tarhini, K. Hone, X. Liu	189

Expanding Database Keyword Search for Database Exploration B. Balla, Z. Chen	198
Exploring Database Keyword Search for Association Studies between Genetic Variants and Diseases D. Verma, H.H. Ali, Z. Chen	206
Mobile Cloud Computing Network Attack and Defense Learning System based on Fuzzy Soft Sets Y. Wang, Z. Liu, Z. Du, Y. Huang	214
Design of the Data Acquisition System based on STM32 H.-f. Zhang, W. Kang	222
A Space Increased Reversible Information Hiding Technique by Reducing Redundant Recording J. Chen, T.-S. Chen, W. Hong, C.-N. Lin, M.-C. Wu, H.-Y. Wu	229
Systematic Approach to Successful Implementation of ITIL N. Ahmad, Z.M. Shamsudin	237
Research on Supply Chain Finance Pricing Problem under Random Demand and Permissible Delay in Payment M. Yang	245
The Market Status Tendency, the Corporation Specified Information, and the Market Announcement Effect of M&A J. Yu, Y. Xu, P. Sun, H. Yu, B. Xu	258
In Contango Versus Backwardation, The Truth may not be in Convenience: Disequilibrium States and the Spot-Forward Balance in Commodity Markets R.J. Bowden, P.N. Posch	266
Research on The Multiple Combination Weight Based on Rough Set and Clustering Analysis – The Knowledge Transfer Risk in IT Outsourcing Taken as An Example L. Yuan, F. Xu	274
The Challenges Facing E-book Publishing Industry in Taiwan C.-C. Lin, W.-C. Chiou, S.-S. Huang	282
The Quality Measurement of China High-Technology Exports J. Xiong, S. Qureshi	290
Modeling Maintenance Effect with Failure-Counting Y. Zhou, G. Kou, D. Ergu	298
Using Fuzzy Integral Approach to Enhance Site Selection Assessment – A Case Study of the Optoelectronics Industry Y.-C. Kuo, S.-T. Lu, G.-H. Tzeng, Y.-C. Lin, Y.-S. Huang	306
An Analysis of Internal Control Problem of State-owned Highway Construction Enterprise—take Huanyu Highway Construction & Development Co., Ltd. as an Example H. Zhou, H. Zhou, Z. Tian	314
Criteria Interactions in Multiple Criteria Decision Aiding: A Choquet Formulation for the TODIM Method L.F.A.M. Gomes, M.A.S. Machado, F.F. da Costa, L.A.D. Rangel	324
Dealing the Selection of Project Management through Hybrid Model of Verbal Decision Analysis P.R. Pinheiro, T.C.S. Machado, I. Tamanini	332
Developing Sustainable SCM Evaluation Model using Fuzzy AHP in Publishing Industry M. Shaverdi, M.R. Heshmati, E. Eskandaripour, A.A.A. Tabar	340
Strategic Vender Selection Criteria C.-Y. Shen, K.-T. Yu	350
Robust Optimization for Multiobjective Programming Problems with Imprecise Information F. Hassanzadeh, H. Nemati, M. Sun	357
An Integrated Inventory Model with Order-Size-Dependent Trade Credit and Quality Improvement C.-J. Chuang, C.-H. Ho, L.-Y. Ouyang, C.-W. Wu	365
Smart Inventory Management System of Food-Processing-and- Distribution Industry C.-C. Liang	373
Fuzzy Flexible Delivery and pickup Problem with Time Windows Y.-Y. Chen	379
Multipath Routing Network Management Protocol for Resilient and Energy Efficient Wireless Sensor Networks C. Velásquez-Villada, Y. Donoso	387
Energy Load Balancing Strategy to Extend Lifetime in Wireless Sensor Networks G.A. Montoya, Y. Donoso	395
Harmonic Fractals in the Brain: Transient Tuning and Synchronic Coordination in the Quasi-Chaotic Background of Ongoing Neural EEG Activity H.A. Díaz, F.M. Córdova	403
Olive Fruits Recognition using Neural Networks G. Gatica, S. Best, J. Ceroni, G. Lefranc	412
A Fast Multifractal Model for Self-Similar Traffic Flows in High-Speed Computer Networks G. Millán, G. Lefranc	420
Goal Programming as a Decision Model for Performance-based Budgeting L. Zamfirescu, C.-B. Zamfirescu	426

A New Approach to Detect and Extract Characters from off-line Printed Images and Text A. Choudhary, R. Rishi, S. Ahlawat	434
Toward a Hybrid Approach of Primitive Cognitive Network Process and Particle Swarm Optimization Neural Network for Forecasting G. Zhang, K.K.F. Yuen	441
Prioritized Multi-Criteria Decision Making based on the Idea of PROMETHEE X. Yu, Z. Xu, Y. Ma	449
Studies on Consistency Measure of Hesitant Fuzzy Preference Relations B. Zhu	457
Study on Interval Intuitionistic Fuzzy Multi-Attribute Group Decision Making Method based on Choquet Integral J. Qin, X. Liu	465
Consensus Rules with Minimum Adjustments for Multiple Attribute Group Decision Making B. Zhang, Y. Dong	473
Optimal Weight Determination and Consensus Formation under Fuzzy Linguistic Environment Y. Xu, H. Wang	482
An Extended LINMAP Method for Multi-Attribute Group Decision Making under Interval-Valued Intuitionistic Fuzzy Environment W. Wang, X. Liu	490
Linguistic Consensus Models based on a Fuzzy Ontology I.J. Pérez, R. Wikström, J. Mezei, C. Carlsson, K. Anaya, E. Herrera-Viedma	498
An EMM-based Approach for Text Classification J.G. Liang, X.F. Zhou, P. Liu, L. Guo, S. Bai	506
Exploring Determinants of Inflation in China based on L_1 -E-twin Support Vector Regression Y.-F. Ye, H. Cao, L. Bai, Z. Wang, Y.-H. Shao	514
Bit-Parallel Multiple Approximate String Matching based on GPU K. Xu, W. Cui, Y. Hu, L. Guo	523
A Multi-Level Analysis Framework in Network Security Situation Awareness H. Zhang, J. Shi, X. Chen	530
DEM: Direct Estimation Method for Photovoltaic Maximum Power Point Tracking J. Ma, K.L. Man, T.O. Ting, N. Zhang, S.-U. Guan, P.W.H. Wong	537
A Middleware Architecture for Price Comparison Service on Mobile Phones G. Li, Z. Shen	545
Study on Government Procurement Behavior based on the Agent Simulation X. Enquan, C. Mengying, L. Cheng	554
State-of-Art Power Line Communications Channel Modelling W. Zhu, X. Zhu, E. Lim, Y. Huang	563
Barriers of B2B e-Business Adoption in Indonesian SMEs: A Literature Analysis I. Janita, W.K. Chong	571
Porting LooCI Components into Zigduino Z. Shen, K.L. Man, H.-N. Liang, N. Zhang, D.O. Afolabi, E.G. Lim	579
Developing a Smart Learning Environment in Universities Via Cyber-Physical Systems C.-U. Lei, K. Wan, K.L. Man	583
Investigation of Neural Networks for Function Approximation S. Yang, T.O. Ting, K.L. Man, S.-U. Guan	586
Analysis on Overlapped and Non-overlapped Data S. Lee, Y. Sun, H. Wei	595
Can Representation Bias help the Returns Forecast and Portfolio Selection? D. Zhao, Y. Fang	603
A Factor Model for the Calculation of Portfolio Credit VaR P. Li, X. Wang, H. Wang	611
Copula based Change Point Detection for Financial Contagion in Chinese Banking X. Zhu, Y. Li, C. Liang, J. Chen, D. Wu	619
The Systematic Controllability of Cotton Production of Xinjiang based-on Control Theory Y. Chen, H. Yang, W. Liu	627
Predicting the Removal of Special Treatment or Delisting Risk Warning for Listed Company in China with Adaboost L. Zhou	633
The Impact of Interest Rate on Information Flow Interpretation: Evidence from ChiNext Y. Zhang, L. Feng, X. Jin, D. Shen	641
The Research of Liquidity Risk Measurements in China Stock Market L. Li, Y. Feng	647
Integrating Credit and Market Risk: A Factor Copula based Method C. Liang, X. Zhu, Y. Li, X. Sun, J. Chen, J. Li	656
The Impact of Investors' Risk Attitudes on Skewness of return Distribution F. Wen, M. Tao, Z. He, X. Chen	664

Method of Value-at-Risk and Empirical Research for Shanghai Stock Market Q. Chen, R. Chen	671
Financial Distress Prediction for Chinese Listed Manufacturing Companies Y. Chen, L. Zhang, L. Zhang	678
Pattern Analysis in the Study of Science, Education and Innovative Activity in Russian Regions F. Aleskerov, L. Egorova, L. Gokhberg, A. Myachin, G. Sagieva	687
Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning L. Zhang, L. Zhang, W. Teng, Y. Chen	695
An Approach to Build Industry Asset Price Index: Take Four Industries for examples F. Meng, Y. Wang, J. Li	704
Data, Information, Knowledge, Wisdom and Meta-Synthesis of Wisdom-Comment on Wisdom Global and Wisdom Cities G. Jifa	713
Scenario Simulation and Policy Analysis on Energy Development in Qinghai Province Y. Yang, X. Sun, X. Zhu, Y. Xie	720
Recent Research in Project Finance - A Commented Bibliography D. Kayser	729
Research of the Influence of Macro-Economic Factors on the Price of Gold L. Lili, D. Chengmei	737
Research on the Contribution of Structure Adjustment on Carbon Dioxide Emissions Reduction based on LMDI Method L. Wu, W. Zeng	744
The Influence Factor Analysis of Comprehensive Energy Consumption in Manufacturing Enterprises G.K.F. Tso, F. Liu, K. Liu	752
Energy Time Series Data Analysis based on a Novel Integrated Data Characteristic Testing Approach L. Tang, C. Wang, S. Wang	759
The Analysis of Peasant Households' Collective Brand Maintenance Behavior in the "One Village, One Product" Program of China P. Xujin, J. Li	770
A Recommendation-based Matchmaking Scheme for Multiple Mobile Social Networks against Private Data Leakage Y. Wang, J. Hou, Y.-w. Tan, X. Nie	781
Mining Explainable User Interests from Scalable user Behavior Data L. Jun, Z. Peng	789
The Research of Weighted Community Partition based on SimHash L. Yang, S. Ying, S. Jixi, X. Kai	797
Advertisement Click-Through Rate Prediction using Multiple Criteria Linear Programming Regression Model F. Wang, W. Suphamitmongkol, B. Wang	803
EPLogCleaner: Improving Data Quality of Enterprise Proxy Logs for Efficient Web usage Mining H. Sha, T. Liu, P. Qin, Y. Sun, Q. Liu	812
Distributed OSN Crawling System based on Ajax Simulation S. Jixi, S. Ying, L. Yang, X. Kai, G. Li	819
Data Stolen Trojan Detection based on Network Behaviors Y. Pu, X. Chen, X. Cui, J. Shi, L. Guo, C. Qi	828
Towards Analyzing and Improving Service Accessibility under Resource Enumeration Attack X. Wang, J. Shi, L. Guo	836
A Multiple Watermarking Algorithm for Texts Mixed Chinese and English X. Rui, C. XiaoJun, S. Jinqiao	844
A Bigram based Real Time DNS Tunnel Detection Approach C. Qi, X. Chen, C. Xu, J. Shi, P. Liu	852
Detecting Sybil Nodes in Anonymous Communication Systems W. Zang, P. Zhang, X. Wang, J. Shi, L. Guo	861
A Continuous Re-Authentication Approach Using Ensemble Learning C. Xiaojun, W. Zicheng, P. Yiguo, S. Jinqiao	870
A Framework of Evaluation Location Privacy in Mobile Network C. Lee, Y. Guo, L. Yin	879
A Three-Dimensional Evaluation Model for Network Management System F. Xu, J. Shi, X. Chen, K.P. Chow	888
Risk Incentive Problems between a Lender & a Borrower and Possible Ways for Mitigation H. Inoue, M. Miyake, S. Takahashi	893
Study on the Influence Factors of High-Tech Enterprise Credit Risk: Empirical Evidence from China's Listed Companies M. Zhang, Y. He, Z.-f. Zhou	901
The Model and Empirical Research of Application Scoring based on Data Mining Methods L. Hui, S. Li, Z. Zongfang	911
Influence of Psychological and Emotional Factors on the Venture Enterprise Value and the Investment Decision-Making G. Jing, C. Hao, Z. Xian	919

The Evolutionary Game Analysis of Credit Behavior of SME in Guaranteed Loans Organization X. Chao, Z. Zongfang	930
Are there Some Manipulations of SHIBOR: A Hypothesis Testing based on Financial Products Linked to SHIBOR L. Chen, Z. Zhou	939
The Research on Credit Risk of Business Groups based on Related Guarantee Y. Yang, L. Li, Z. Zongfang	945
An Agent-based Model of Social Norm and Cooperation T. Yu, L. Zhang	951
An Evolutionary Game Model of Financial Markets with Heterogeneous Players H. Li, C. Wu, M. Yuan	958
Vulnerability Analysis of Symbiosis Networks of Industrial Ecology Parks Y. Zeng, R. Xiao, X. Li	965
Evolutionary Mechanism and Information Supervision of Public Opinions in Internet Emergency D. Liu, W. Wang, H. Li	973
Structure Emergence in the Evolution of Social Networks and its Case Study Z. Tao, R. Xiao, L. Wang	981
Research on Delay Characteristics of Information in Scale-Free Networks based on Multi-Agent Simulation X. Gong, J. Xu	989
Performance Evaluation of a Supply Chain Network T. Chen, X. Gong	1003
Common Asset Holdings and Systemic Risk of Financial Network Z. Zhao, W. Zhang, S. Shi	1010
Influence Analysis of FDI on China's Industrial Structure Optimization Z. Qiong, N. Minyu	1015
Sensitivity Analysis of b in MCLP Model for Credit Card-Holders Classification A. Li, Y. Yu, Z. Liu	1023
The Best Separating Decision Tree Twin Support Vector Machine for Multi-class Classification Y.-H. Shao, W.-J. Chen, W.-B. Huang, Z.-M. Yang, N.-Y. Deng	1032
Feature Selection based on Linear Twin Support Vector Machines Z.-M. Yang, J.-Y. He, Y.-H. Shao	1039
An Embedded Backward Feature Selection Method for MCLP Classification Algorithm M. Zhu, J. Song	1047
Support Vector Regression for Newspaper/Magazine Sales Forecasting X. Yu, Z. Qi, Y. Zhao	1055
Nonparallel Support Vector Machines for Multiple-instance Learning Q. Zhang, Y. Tian, D. Liu	1063
A Multiple-Criteria and Multiple-Constraint Levels Linear Programming based Error Correction Classification Model B. Wang, Y. Wang	1073
Knowledge-based Support Vector Classification based on C-SVC C. Zhang, X. Shao, D. Li	1083
Toward a Ranking Strategy for e-commerce Products in an e-alliance Portal using Primitive Cognitive Network Process K.K.F. Yuen	1091
IT for Corruption Suppression A. Vasin, P. Nikolaev, A. Urazov	1097
The Threshold Decision Making F.T. Aleskerov, V.V. Chistyakov	1103
Potential Optimality of Pareto Optima V.V. Podinovski	1107
Estimation of Conflict and Decreasing of Ignorance in Dempster-Shafer Theory A. Lepskiy	1113
Super-threshold Procedures and Their Application to the Search Problem F. Aleskerov, E. Mitichkin, S. Shvydun, V. Yakuba	1121
Dry Bulk Time Charter Rates Joint Return Distribution Modeling: Copula-Approach A. Merikas, A. Merika, H. Penikas	1125
Matrices of Forests, Analysis of Networks, and Ranking Problems P. Chebotarev, R. Agaev	1134
Three Quantitative Management Problems in Public Procurement and Decision Procedures for their Analysis and Solving A.S. Belenky	1142
A New Data Processing Algorithm based on Model Fitting for Wireless Sensor Networks G. Huang, Y. Chen, Y. Zhou, P. Liu, S. Liu	1154
A Task Scheduling Algorithm based on QoS-Driven in Cloud Computing X. Wu, M. Deng, R. Zhang, B. Zeng, S. Zhou	1162

A Trust Evaluation Model for Cloud Computing X. Wu, R. Zhang, B. Zeng, S. Zhou	1170
Customer Satisfaction Scoring for Innovative Product Management based on Extenics Z. Xiang, X. Li, J. Yang	1178
Extenics-based Test Case Generation for UML Activity Diagram L. Li, X. Li, T. He, J. Xiong	1186
Study on Innovation Capability of College Students based on Extenics and Theory of Creativity A. Chen, L. Li, X. Li, J. Zhang, L. Dong	1194
Study on Business Process Knowledge Creation and Optimization in Modern Manufacturing Enterprises L. Shu, S. Liu, L. Li	1202
The Study of the Relationship between Interaction Orientation and Business Performance P. Yuan, Y. Liu	1209
Extensible Multi-criteria Optimization Classifier for Prediction of Chinese Semantic Word-formation Patterns Z. Zhang, X. Li, H. Zou, S. Kang, G. Gao	1216
Analysis of the Badminton Match-throwing in London Olympics H. Yang, W. Liu	1222
A New Measurement of Sectoral Concentration of Credit Portfolios Y. Chen, X. Wei, L. Zhang	1231
Sector Indices Correlation Analysis in China's Stock Market D. Cao, W. Long, W. Yang	1241
Trade Duration and Liquidity of Chinese Stock Market B. Li, M. Qi	1250
A Comparative Study of Determinants of International Capital Flows to Asian and Latin American Emerging Countries H. Yang, Y. Xiong, Y. Ze	1258
Investigation and Analysis on Current Situation of Rural Cooperative Finance in Xinjiang J. Ismail, W. Xianhua	1266
The Probe Algorithm with Information Map: A New Dynamic Method for Correlation Analysis Y. Liu, X. Huang, K. Guo	1276

Information Technology and Quantitative Management (ITQM2013)

Super-threshold Procedures and Their Application to the Search Problem

Fuad Aleskerov^{a,b}, Evgeny Mitichkin^c, Sergey Shvydun^a *, Vyacheslav Yakuba^b

^aNational Research University Higher School of Economics, 11 Pokrovsky boulevard, Moscow 109028, Russia

^bInstitute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznaya street, Moscow 117997, Russia

^cUniversity of Mannheim, Schloss, Mannheim 68131, Germany

Abstract

The paper examines a choice problem in case of large number of alternatives characterized by a set of criteria. Very often existing choice procedures cannot be used due to their computational complexity. In contrast, the approach proposed in this paper utilizes a set of easy to use techniques having complexity less than the quadratic. We apply super-threshold procedures sequentially in order to find the best alternatives. The application to learning to rank problem is also considered. Finally, two methods of defining thresholds and order of criteria in super-threshold procedures are studied. While the first method is based on the etalon criterion, the second method uses the distribution function. Both methods developed are mainly oriented on the choice of search results in search engines and capable of processing large amount of data in networks, but can also be applied to any system in which the alternatives are represented by a set of criteria.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management

Keywords: Super-threshold Choice Procedures; Search Problem, Learning to Rank; Adaptive Filtering.

1. Introduction

Consider a finite set A of alternatives evaluated by n criteria, i.e. the vector of criterial values (x_1, \dots, x_n) is assigned to each alternative x from A

$$x \in A \rightarrow (x_1, \dots, x_n).$$

The problem lies in constructing a transformation φ — the rule of aggregation over A — such that $\varphi: A \rightarrow \mathbb{R}^1$.

There are a lot of choice procedures applicable to cases with relatively small number of alternatives and criteria. However, in case of thousands of alternatives and criteria, existing procedures cannot provide sufficient results in a reasonable time. We propose to use super-threshold procedures for the choice of alternatives.

Super-threshold procedures are based on the rule that chooses all alternatives for which the estimate criterion value is greater or equal than some assigned threshold [1]. The main idea of this approach is to consider the consistent application of super-threshold procedures to different criteria in order to find the best alternatives. Comparing to other procedures the computational complexity of super-threshold procedures is lower (normally lower than quadratic) as every iteration decreases the number of alternatives.

The choice problem can be divided into two types. In the first type the information about the utility of alternatives is not specified. The utility of the alternative denotes how well this alternative meets the need of the end user. The utility of alternative can vary: the alternative with a high utility value is considered as the best alternative (choice) while alternative with a low utility value is considered as a bad alternative. In this case, the threshold values and the order of elimination are normally defined manually.

* Corresponding author. Tel.: +7-915-218-6094.

E-mail address: shvydun@hse.ru.

The second type is the learning to rank problem which is based on the automatic construction of choice models using training dataset. The choice model is a set of rules that is used to choose the best alternatives. The training data consists of lists of items with partial order specified between items in each list. This order is typically induced by a numerical or ordinal score or a binary judgment (e.g. "relevant" or "not relevant") for each item. Choice model's purpose is to choose (select) items in a new, unseen list in a way, which is similar to the choice in the training data in some sense. In this case the threshold level as well as the order of procedures should be defined automatically.

In this paper, we propose two methods of how the order of criteria and the threshold value for each criterion can be defined. While the first method is based on the use of etalon criterion, the second method is based on the distribution function. Both methods developed are mainly oriented on the choice of search results in search engines and capable of processing large amount of data in networks, but can also be applied to any system in which the objects are represented by a set of criteria.

2. Super-threshold procedures

The main principle of the super-threshold procedures incorporates the sequential application of filters to the data, in order to build a system of constraints which is capable to automatically distinguish alternatives with a high utility values and other alternatives according to their criterial values.

Two methods of defining the order of elimination and threshold values are proposed

- Super-threshold etalon-based procedures,
- Super-threshold procedures based on the distribution function.

2.1. Super-threshold etalon-based procedures

All alternatives can be represented as points in (n+1)-dimensional space where n dimensions are criterial values of alternatives and one dimension is the utility. The procedure is divided into two steps. The first step of the procedure is the selection of criterion with the densest distribution of high-utility alternatives, i.e. the selection of criterion which criterion values of high-utility alternatives are closest to each other.

$$\varphi^* = \{\varphi_i \in \varphi \mid \forall j, j \neq i: \max_{x \in A^*} \varphi_i(x) - \min_{x \in A^*} \varphi_i(x) \leq \max_{x \in A^*} \varphi_j(x) - \min_{x \in A^*} \varphi_j(x)\},$$

where φ is the set of criteria, A^* is the set of the best (high-utility) alternatives, $A^* \subset A$.

The chosen criterion is considered as the etalon. The direction of the super-threshold elimination depends on whether it is necessary to minimize or maximize the criterion. The threshold value can be either preset initially (mean, median, or other fixed value) or can be chosen or adjusted automatically.

The second step consists of construction of the order of the eliminations in order to increase the percentage of high-utility alternatives on the etalon criterion.

The order of elimination can be accomplished according to the following rule: the super-threshold choice is performed on the criterion (except the etalon) that excludes as much bad alternatives as possible within the accumulation (interval) of the best alternatives on the etalon criterion. When the criterion for elimination is found, it is included to the order of elimination and excluded from further consideration; alternatives, which values are below the threshold, are excluded from the set A and the procedure (recalculation the threshold and the order steps) iterates again. Thus, the procedure continues until the goal is achieved (only alternatives with a high utility value remain) or all criteria are used.

2.2. Super-threshold procedures based on the distribution function

All alternatives are divided into two groups – alternatives with high utility value and alternatives with low utility value (all other alternatives). Each group is presented by a set of distribution functions based on all criteria. In this case the distribution function describes the ratio of alternatives A for which the criterion value is less than or equal to value x .

The threshold value for each criterion is defined as the criterion value where the difference V^* between two

distribution functions reaches its maximum.

$$V^* = \max_{x \in A} |F_1(x) - F_2(x)|,$$

where $F_1(x)$, $F_2(x)$ are the values of distribution functions depending on x .

It can be inferred that the concentration of alternatives with a high utility value among the others is highest at the point where the criterion value equals to x .

The order of elimination can be accomplished according to the following rule: the super-threshold choice is performed on the criterion with the highest difference V^* between distribution function of high-utility alternatives and the distribution function of other alternatives. When the criterion for elimination is found, it is included to the order of elimination and excluded from further consideration; alternatives, the values of which are below the assigned threshold, are excluded and the procedure (recalculation of the threshold etc.) iterates again. As before, the procedure continues until the goal is achieved or all criteria are used.

As in etalon-based procedure, the result of the procedure based on the distribution function is the sequence of criteria accompanied with threshold values for the corresponding criteria (choice rule) that is used to define or predict the utility of new alternatives.

3. Application of super-threshold procedures to the search problem

A set from the LETOR 4.0 [2] collection was chosen to test two above-described methods. LETOR 4.0 is a special dataset with 10000 of queries provided by Microsoft Research to cope with the problem of learning to rank. It contains a set of evaluated web pages (documents). Each document is characterized by a set of criteria and has a relevance value. In information science and information retrieval, the relevance denotes how well a retrieved document or set of documents meets the information need of the user. The relevance judgments are obtained from a set of a commercial web search engine (Microsoft Bing), which take five values from 0 (irrelevant) to 4 (perfectly relevant). A query-document pair is represented by a 136-dimensional feature vector.

After a series of preliminary experiments some important properties of the set were noted, namely:

- Different queries and sets of documents are represented by different key features
- Maximization or minimization algorithms applied to the whole dataset do not choose (select) only the best documents with a high accuracy.
- It was noticed that simultaneous application of super-threshold procedures on several key (significant) features can lead to the required result.

Thus, the methods were tested on a subset of queries with the same choice rules. The testing was performed with the use of key (significant) features that were found by some procedures of data analysis such as the data transformation and the dimension reduction [3]. Key (significant) features are features that most affect the choice, so their existence is essential to the construction of the choice model. The main task of both methods was to estimate their average accuracy that can be defined as the ratio of the best alternatives in the remaining subset. In other words, the average accuracy shows how well both methods can distinguish relevant and irrelevant documents according to their feature's values. The average accuracy of the first method on the chosen set is 65%; the average accuracy of the second method is 46%. The first method showed better result on small dataset, the second demonstrated higher results in the case with different choice rules.

The computational complexity of both methods is also important. Both algorithms have the complexity that is less than quadratic, so the increase of the number of alternatives does not influence so much the computational complexity of both methods.

4. Conclusion

Super-threshold procedures are proposed to be used for the choice problem with large number of alternatives characterized by a set of criteria. One of the key advantages of the procedures is that the computational complexity is less than quadratic as every iteration decreases the number of alternatives. Two methods of defining the threshold and the elimination order in learning to rank problem are proposed. The first method is based on the use of the etalon criteria; the second method relies on distribution functions. Both procedures were tested on the LETOR 4.0

dataset. The experiments showed that the goal to develop self-learning choice model for the choice of alternatives is achieved.

References

1. M. Aizerman, F. Aleskerov F., Theory of Choice, Elsevier, North-Holland, 1995.
2. H. Li, T.Y. Liu, T. Qin, J. Xu, LETOR: A Benchmark Collection for Research on Learning to Rank for Information Retrieval, Information Retrieval Journal, Vol. 13, No. 4, 346-374, 2009.
3. E.M. Braverman, I.B. Muchnik, The Structural Methods for Empirical Data Processing, NAUKA Publishers, Moscow, 1983.