Advances in Intelligent Systems and Computing 1030

Srikanta Patnaik
Andrew W. H. Ip
Madjid Tavana
Vipul Jain *Editors*

New Paradigm in Decision Science and Management

Proceedings of ICDSM 2018

Springer

Advances in Intelligent Systems and Computing

Volume 1030

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Srikanta Patnaik · Andrew W. H. Ip ·
Madjid Tavana · Vipul Jain
Editors

New Paradigm in Decision Science and Management

Proceedings of ICDSM 2018

Springer

Editors
Srikanta Patnaik
Siksha 'O' Anusandhan University
Bhubaneswar, India

Madjid Tavana
Department of Business Systems
and Analytics
La Salle University
Philadelphia, PA, USA

Andrew W. H. Ip
Department of Industrial
and Systems Engineering
The Hong Kong Polytechnic University
Kowloon, Hong Kong

Vipul Jain
Victoria Business School,
School of Management
Victoria University of Wellington
Wellington, New Zealand

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-13-9329-7 ISBN 978-981-13-9330-3 (eBook)
https://doi.org/10.1007/978-981-13-9330-3

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The International Conference on Decision Science and Management (ICDSM-2018) was organized by the Interscience Institute of Management and Technology (IIMT), Bhubaneswar, India, on December 22–23, 2018. ICDSM-2018 has provided an ideal platform for researchers, academicians, software developers, and industry experts to share their research works, findings, and new research directions with each other. The conference focuses on research and developmental activities in the field of decision science and its significance in various fields of management. A good response has been received for this conference, and the selected papers out of the total number of submissions have been divided into three major parts, i.e., data science and analytics, decision science and management, and finally ICT and innovative applications.

The first part, data science and analytics, is an interdisciplinary field as it provides a strong combination of algorithms with mathematics and statistics. The underlying techniques are capable of gathering a large amount of data from different sources and working with huge unstructured datasets and apply efficient techniques to extract meaningful, effective, and accurate insights using data mining, neural networks, and machine learning techniques. These insights further help in identifying hidden patterns and future trends and making actionable decisions for solving significant problems and create impact. Data science and analytics when combined with business management concepts can be used to solve some of the most crucial problems. Many of the submissions fall into this section of the proceeding.

The second part is decision science and management. Decision science is a multidisciplinary area that involves existing as well as emerging fields such as computer science, information technology, mathematics, statistics, and business management for solving several crucial problems of management processes. Various management processes and systems are being used to process data collected from heterogeneous sources for extracting information and provide significant insights. These insights support managers for enhancing their understanding of important problems in hand and make crucial decisions in uncertain and critical situations. Decision science thus adds value to various processes of business management sectors such as evaluation process while hiring new candidates to assist senior managements in dealing with vulnerable circumstances by keeping track of influential changes occurring in other sectors. Further, it provides new direction to the advancement in key areas such as adoption of Internet of things, distributed computing, human resource developments, resource management, marketing strategies, supply chain management, logistics, business operations, financial markets, thus leading the integration of data science and decision science techniques in business sector, a potential research area. We have received a good number of submissions in this section.

The third part is ICT and innovative applications which explore various application areas of industry as well as other sectors that utilize both data science and decision science for solving problems. Information and communication technology (ICT) is an extension of information technology and involves all the infrastructure and components that enable modern computing including computer systems, mobile devices, networking components, and applications. It allows organizations (including business firms to government bodies) and individuals to interact with each other in the digital world. This section of the conference has also received lots of attention. The conference not only encouraged the submission of unpublished original articles in the fields of data science, decision science, and management issues but also considered the cutting-edge applications in organizations and firms while scrutinizing the relevant papers.

Kowloon, Hong Kong — Andrew W. H. Ip
Philadelphia, USA — Madjid Tavana
Bhubaneswar, India — Srikanta Patnaik
Wellington, New Zealand — Vipul Jain

Acknowledgements

The contributions covered in this proceeding are the outcome of the contributions from more than one hundred researchers. We are thankful to the authors, paper contributors of this volume, and the departments which supported the event a lot.

We are thankful to the Editor-in-Chief of the Springer Book series on *Advances in Intelligent Systems and Computing* Prof. Janusz Kacprzyk for his support to bring out the maiden volume of the conference, i.e., ICDSM-2018.

We are also very much thankful to Dr. Aninda Bose, Senior Editor, Hard Sciences, Springer Nature publishing, and his team for his constant encouragement support and time-to-time monitoring.

We express our deep sense of gratitude to Prof. Taosheng Wang, Dean, School of Business, Hunan International Economics University, China, for addressing as the guest of honor coming from China, and Prof. Rabi Narayan Subudhi, School of Management, KIIT, for addressing the conference as Chief Speaker.

We would like to record our thanks to our academic partner Hunan International Economics University, China, who will be hosting the next edition of the conference. Last but not least, we are thankful to Interscience Institute of Management and Technology, Bhubaneswar, India, for hosting the event successfully.

We are sure that the readers shall get immense benefit and knowledge from this edited volume. We look forward to your valuable contribution and support to the next editions of ICDSM, which will be held in Hunan International Economics University, Changsha, China.

Andrew W. H. Ip
Madjid Tavana
Srikanta Patnaik
Vipul Jain

Contents

About the Editors

Dr. Srikanta Patnaik is a Professor at the Department of Computer Science and Engineering, Faculty of Engineering and Technology, SOA University, Bhubaneswar, India. Dr. Patnaik has published more than 60 research papers and book chapters, 2 textbooks, and has edited 12 books. He is the Editor-in-Chief of the International Journal of Information and Communication Technology and International Journal of Computational Vision and Robotics, both published by Inderscience, England, and Editor-in-Chief of the book series "Modeling and Optimization in Science and Technology" published by Springer, Germany.

Dr. Andrew W. H. Ip received his PhD from Loughborough University (UK), MBA from Brunel University (UK), MSc in Industrial Engineering from Cranfield University (UK) and LLB (Hons) from the University of Wolverhampton (UK). Dr. Ip is Principal Research Fellow of the Hong Kong Polytechnic University, Professor Emeritus and Adjunct Professor of Mechanical Engineering, University of Saskatchewan, Canada. He is a Visiting Professor at the Civil Aviation University of China, South China Normal University, and University of Electronic Science and Technology of China, and Honorary Industrial Fellow of the University of Warwick, Warwick Manufacturing Group. He is editor in chief of Enterprise Information Systems, Taylor and Francis (SCI indexed); founder and editor in chief of International Journal of Engineering Business Management, SAGE (ESCI and Scopus indexed).

Dr. Madjid Tavana is a Professor and Distinguished Chair of Business Analytics at La Salle University in Philadelphia, where he serves as Chairman of the Business Systems and Analytics Department. He also holds an Honorary Professorship in Business Information Systems at the University of Paderborn in Germany. Dr. Tavana is a Distinguished Research Fellow at the Kennedy Space Center, the Johnson Space Center, the Naval Research Laboratory at Stennis Space Center, and the Air Force Research Laboratory. He has published 15 books and over 250 research papers in scholarly academic journals, and is the Editor-in-Chief of several international journals.

Dr. Vipul Jain is currently working in the area of Operations and Supply Chain Management at Victoria Business School, Victoria University of Wellington, New Zealand. He has also worked as a French Government researcher for the French National Institute for Research in Computer Science and Control at Nancy, France. He is the Editor-in-Chief of the International Journal of Intelligent Enterprise and an Editorial Board Member for seven international journals.

Data Science and Analytics

Quality Index Evaluation Model Based on Index Screening Model

Fenglin Zhang, Mengpei Guo and Xingming Gao

Abstract The principal component analysis theory is quite mature in the simplification and comprehensive evaluation of real data, and the traditional principal component analysis only strives for the form that the index value is real or interval number. It lacks certain persuasion in the face of large uncertainty index values, but it lacks the simplification of the three-parameter interval number. Based on the principal component analysis theory that the index data is an interval number, the paper proposed the M-PCA (three-parameter interval principal component analysis) method to screen out the unrelated indicators. Principal component analysis does not consider the influence of the negative factor when it is used to reduce the dimension of the comprehensive index. There are some misunderstandings and defects in the application. In dealing with the weights of unrelated indicators, this paper considered the subjective theories to solve the weights of the new indicators, combined with the multi-attribute gray target decision model to evaluate quality indicators. Finally, it combined with examples and verified that the method is more scientific, reasonable, and effective.

Keywords Principal component analysis · Three-parameter interval · The weights of new indicators · Gray target decision · Quality indicators

---

F. Zhang (✉) · M. Guo · X. Gao (✉)
College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Road, Nanjing 211106, Jiangsu, China
e-mail: zflnuaa@163.com

X. Gao
e-mail: 1136624944@qq.com

M. Guo
e-mail: 772028639@qq.com

S. Patnaik et al. (eds.), *New Paradigm in Decision Science and Management*, Advances in Intelligent Systems and Computing 1030,
https://doi.org/10.1007/978-981-13-9330-3_1

1 Introduction

Quality is the lifeline of an enterprise. The quality of products is decisive for the development of enterprises. Nowadays, a reasonable evaluation of quality indicators can help companies more scientifically understand their current level of quality. How to evaluate the form of the quality index value as the three-end interval number will be a difficult problem for researchers to study in different environments and conditions. To evaluate the problem, the comprehensive evaluation of the indicator is an important link, and the scientific and reasonable method has a greater impact on the final result. Multi-attribute evaluation is a systematic analysis method for studying the problem of uncertainty evaluation. It is the problem of selecting, ranking, or evaluating multiple events and multiple countermeasures. Its purpose is to improve the decision-making process and systematically evaluate and analyze quality indicators. Multi-attribute evaluation methods such as analytic hierarchy process, TOPSIS method, and fuzzy decision-making method have attracted the attention of many scholars since their introduction. However, the above studies directly deal with the original indicator values, ignoring the interrelationship between the various quality indicators. The degree of preprocessing of the evaluation index is related to the quality and weakness of the quality evaluation and the complexity of the calculation, especially for the case where the quality index value is not a real number, such as the number of intervals or the number of three-parameter intervals.

Many scholars have conducted in-depth research on the interrelationship between evaluation indicators. Ait-Izem et al. [1] compared the four most common interval principal component analysis (PCA) methods and applied them to fault detection and isolation. Based on the reconstruction principle in the classical PCA method, the reconstruction of the interval is optimized, and a new criterion is derived to determine the structure of the interval PCA model. Liu et al. [2] proposed a new model. This new model is an interval-intuitive fuzzy principal component analysis. Based on the traditional principal component analysis, it treats decision makers and attributes as interval-intuitive fuzzy variables, and transforms these two kinds of variables through the proposed principal component analysis model. As an independent variable, Yu et al. [3] used principal component analysis to reduce the index of the stock price and predicted the stock price based on the generalized regression neural network model. Hou [4] proposed two improved methods, one is the principal component analysis method based on empirical statistics, and the other is the principal component analysis method based on interval matrix operation, which achieves the expected effect. Qi and Xiong [5] put forward two kinds of energy-saving evaluation methods for the central heating boiler room: qualitative and quantitative index.

In this paper, when processing the simplification of index data for the three-parameter interval data, it is based on the three-interval principal component analysis of the most probable value method (M-PCA—three-parameter interval principal component analysis) to filter out the indicators with low correlation and for the new indicator weight analysis. Finally, the multi-attribute gray target decision model is used to evaluate the quality indicators.

2 Three-Parameter Interval Principal Component Analysis

2.1 Principal Component Analysis

PCA is able to turn multiple indicators into fewer comprehensive indicators. The principle of comprehensive variables is the linear combination of multiple indicators.

2.2 Calculation Steps

(1) Constructing a sample matrix

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{1}$$

(2) Transform sample array X to $Q = \lfloor q_{ij} \rfloor_{m \times n}$

$$q_{ij} = \begin{cases} x_{ij}, & \text{Positive intensity indicator} \\ 1/x_{ij}, & \text{Negative intensity indicator} \end{cases} \tag{2}$$

(3) Normalize matrix Q to obtain a normalized matrix Z

$$Z = \begin{bmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_m^T \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \dots & z_{mn} \end{bmatrix} \tag{3}$$

$$z_{ij} = \frac{q_{ij} - \bar{q}_j}{s_j} \tag{4}$$

The standardization makes the variance of all indexes equal to 1, smoothing the variation of each indicator. The main component of the standardized data extraction contains information on the mutual influence of each index and does not accurately reflect all the information contained in the original data. To eliminate the influence of the dimension and magnitude of variables, this paper uses the “average” method.

$$z_{ij} = \frac{q_{ij}}{\sum_{j=1}^{m} q_{ij}} \tag{5}$$

(4) Calculate the covariance matrix of the averaging matrix z

$$Y = \left[y_{ij}\right]_{\ln n} = \frac{Z^T Z}{m-1} \tag{6}$$

(5) Solve the eigen value

$$|Y - \lambda I_n| = 0 \tag{7}$$

Solve to get n feature values $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$.

(6) Determine the p-value. In general, the utilization rate of information is more than 85%.

$$\frac{\sum_{j=1}^{p} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \geq 0.85 \tag{8}$$

Solving equations: $Yb = \lambda_j b$.

$$b_j^0 = \frac{b_j}{\left|b_j\right|} \tag{9}$$

(7) Find the p principal components of $z_i = (z_{i1}, z_{i2}, \ldots, z_{\text{in}})$

$$u_{ij} = z_{,}^T b_j^0, \; j = 1, 2, \ldots, p \tag{10}$$

Decision matrix:

$$U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \ldots & u_{1p} \\ u_{21} & u_{22} & \ldots & u_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ u_{m1} & u_{m2} & \ldots & u_{mp} \end{bmatrix} \tag{11}$$

2.3 M-PCA Three-Parameter Interval Principal Component Analysis

In the actual decision problem, the interval range is made too large, which will increase the uncertainty of the decision result. It introduced three-parameter interval numbers and added the most likely number. This paper is based on the M-PCA to solve the three-parameter interval index principal component values.

2.4 Calculation Steps

(1) Construction matrix X:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix} = \begin{bmatrix} [x_{11}^S, x_{11}^M, x_{11}^L] & [x_{12}^S, x_{12}^M, x_{12}^L] & \cdots & [x_{1n}^S, x_{1n}^M, x_{1n}^L] \\ [x_{21}^S, x_{21}^M, x_{21}^L] & [x_{22}^S, x_{22}^M, x_{22}^L] & \cdots & [x_{2n}^S, x_{2n}^M, x_{2n}^L] \\ \vdots & \vdots & \vdots & \vdots \\ [x_{m1}^S, x_{m1}^M, x_{m1}^L] & [x_{m2}^S, x_{m2}^M, x_{m2}^L] & \cdots & [x_{mn}^S, x_{mn}^M, x_{mn}^L] \end{bmatrix} \tag{12}$$

where x_{ij}^s represents the minimum value of the j index of the i product, x_{ij}^M represents the most likely value of the j index of the i product, and x_{ij}^L represents the maximum value of the j index of the i product.

(2) Constructing the most likely index value matrix X_M Perform a traditional covariance-based principal component analysis for Eq. (11) to obtain a decision matrix $u_{ij}^M = z_i^M b_j^0$, $j = 1, 2, \ldots, p$.
(3) Solving sample endpoint number decision matrix:

$$X^M = \begin{bmatrix} \left(x_1^M\right)^T \\ \left(x_2^M\right)^T \\ \vdots \\ \left(x_m^M\right)^T \end{bmatrix} = \begin{bmatrix} x_{11}^M & x_{12}^M & \cdots & x_{1n}^M \\ x_{21}^M & x_{22}^M & \cdots & x_{2n}^M \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1}^M & x_{m2}^M & \cdots & x_{mn}^M \end{bmatrix} \tag{13}$$

$$u_{ij}^S = \min_{z_{ij}^S \le z_{ij} \le z_{ij}^L} b_j^0 \tag{14}$$

$$u_{ij}^L = \max_{z_j^S \le z_{ij} \le z_{ij}^L} b_j^0 \tag{15}$$

The linear weighted comprehensive evaluation method is established based on the weighted contribution ratio of the principal components, without considering the influence of negative factors, and there are some misunderstandings and defects in the application.

Because of measurement errors or difficult to measure, this value is not an exact real number and is therefore represented by a three-parameter interval number.

Therefore, it noted that x_{ij} is the index value of the product a under the attribute is b is $x_{ij} \in \left[x_{ij}^s, x_{ij}^M, x_{ij}^L\right]\left(0 \le x_{ij}^S \le x_{ij}^M \le x_{ij}^L,\ i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n\right)$. In order to eliminate the dimension and increase comparability, the data is averaged to obtain a normalized decision matrix $U = \left(u_{ij}\right)_{m\times n}$, where $u_{ij} \in \left(u_{ij}^S, u_{ij}^M, u_{ij}^L\right)$ is the three-parameter interval gray number, which represents the index value of the product a under the attribute b.

Definition 1
$u_j^+(\otimes) = \max\left\{\left(u_{ij}^S + u_{ij}^M + u_{ij}^L\right)/3\right\} i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$ the corresponding effect value is denoted $\left[\left(u_{ij}^S\right)^+, \left(u_{ij}^M\right)^+, \left(u_{ij}^L\right)^+\right]$, and

$$\begin{aligned} u^+(\otimes) =& \left\{u_1^+(\otimes), u_2^+(\otimes), \ldots, u_n^+(\otimes)\right\} \\ =& \left\{\begin{array}{l} \left[\left(u_{i1}^S\right)^+, \left(u_{i1}^M\right)^+, \left(u_{i1}^L\right)^+,\right], \left[\left(u_{i2}^S\right)^+, \left(u_{i2}^M\right)^+, \left(u_{i2}^L\right)^+,\right], \\ \quad \ldots, \left[\left(u_{\text{in}}^S\right)^+, \left(u_{\text{in}}^M\right)^+, \left(u_{\text{in}}^L\right)^+,\right] \end{array}\right\} \end{aligned} \tag{16}$$

is called the optimal ideal effect vector.

2.5 The Determination of Subjective Index Weight Based on Analytic Hierarchy Process

Analytic hierarchy process (AHP) divides decision-making problems into different hierarchical structures and compares them at the same level, finds the weights of the same level, and finally analyzes the problems comprehensively.

2.6 Index Evaluation

Let $u_i = \{u_{i1}(\otimes), u_{i2}(\otimes), \ldots, u_{\text{in}}(\otimes)\}$ is the effect evaluation vector of scheme a_i, and $u^+(\otimes) = \left\{u_1^+(\otimes), u_2^+(\otimes), \ldots, u_n^+(\otimes)\right\} = \left\{\left[\left(u_{i1}^S\right)^+, \left(u_{i1}^M\right)^+, \left(u_{i1}^L\right)^+,\right], \left[\left(u_{i2}^S\right)^+, \left(u_{i2}^M\right)^+, \left(u_{i2}^L\right)^+,\right], \ldots, \left[\left(u_{\text{in}}^S\right)^+, \left(u_{\text{in}}^M\right)^+, \left(u_{\text{in}}^L\right)^+,\right]\right\}$ is the optimal ideal effect vector. We call

$$\varepsilon_i^+ = 3^{-1/2}\left\{\sum_{j=1}^n \omega_j\left[\left(u_{ij}^s - \left(u_{ij}^S\right)^+\right)^2 + \left(u_{ij}^M - \left(u_{ij}^M\right)^+\right)^2 + \left(u_{ij}^L - \left(u_{ij}^L\right)^+\right)^2\right]\right\}^{1/2} \tag{17}$$

is the target center distance of the effect vector of scheme a_i. The smaller ε_i^+ is, the better the a_i scheme is. The bigger ε_i^+ is, the worse the a_i scheme is.

3 Analysis of Examples

When a unit purchases artillery weapons, there are four series of new types of guns a_i available for selection, and consider the following eight attributes: b_1 is the fire assault capacity coefficient; b_2 is the response capability index; b_3 is the mobility index; b_4 is the viability Index; b_5 is cost; b_6 is stability capability index; b_7 is index of risk resistance; and b_8 is an operational index. Recorded as the index value of the product under the attribute, the four types of artillery are new types of artillery. Due to measurement errors or difficulties in measurement, the value is not an exact real number and is therefore represented by a three-parameter interval number.

$$x_{ij} \in \left[x_{ij}^S, x_{ij}^M, x_{ij}^L\right]\left(x_{ij}^S \le x_{ij}^M \le x_{ij}^L, i = 1, 2, \ldots, 4; j = 1, 2, \ldots, 8\right).$$

Each scenario with respect to evaluation attributes is shown in the matrix X.

$$\left[\begin{array}{llll} [26000, 26460, 27000] & [2, 3, 6, 4] & [18000, 18480, 19000] & [0.7, 0.75, 0.8] \\ [60000, 65400, 70000] & [3, 3.4, 4] & [16000, 16780, 19000] & [0.3, 0.36, 0.4] \\ [50000, 57600, 60000] & [2, 2, 7, 3] & [15000, 15360, 16000] & [0.7, 0.74, 0.8] \\ [40000, 44800, 50000] & [1, 1.8, 2] & [28000, 28680, 29000] & [0.4, 0.46, 0.5] \end{array}\right.$$

$$\left.\begin{array}{llll} [15000, 15560, 16000] & [0.4, 0.48, 0.5] & [4, 4.6.6] & [6.5, 7, 7.5] \\ [27000, 27880, 28000] & [0.6, 0.68, 0.8] & [5, 5.8, 6] & [8, 8.5, 9] \\ [24000, 25380, 26000] & [0.6, 0.64, 0.7] & [6, 6.7, 7] & [7, 2, 8, 8] \\ [15000, 15620, 17000] & [0.5, 0.56, 0.6] & [3.6, 4, 4.8] & [6, 6.5, 7] \end{array}\right]$$

$\varepsilon_1^+ = 0.086$, $\varepsilon_2^+ = 0.057$, $\varepsilon_3^+ = 0.103$, $\varepsilon_4^+ = 0.005$. Due to $\varepsilon_4^+ < \varepsilon_2^+ < \varepsilon_1^+ < \varepsilon_3^+$ the force should give priority to the fourth artillery. After a period of use and analysis of the use of other units, the force was found to be superior to other types of weapons. The results show that the proposed theory has a strong application value.

4 Conclusion

A novel M-PCA three-parameter interval principal component analysis method is proposed, which can extract mutually unrelated index attributes. Then, the multi-attribute target decision model is used to evaluate the quality indicators. Combined with examples, this method proved to be more scientific and effective. For the comprehensive consideration of the subjective and objective weights of the new principal component analysis will be the next research direction.

References

1. Ait-Izem, T., Harkat, M.F., Djeghaba, M., et al.: On the application of interval PCA to process monitoring: a robust strategy for sensor FDI with new efficient control statistics. J. Process Control **63**(1), 29–46 (2018)
2. Liu, B., Shen, Y., Zhang, W., et al.: An interval-valued intuitionistic fuzzy principal component analysis model-based method for complex multi-attribute large-group decision-making. Eur. J. Oper. Res. **245**(1), 209–225 (2015)
3. Yu, Z., Qin, W., Zhao, Z., Xiang.: The stock price prediction based on principal component analysis and generalized regression neural network. Stat. Decis. Mak. **34**(18), 168–171 (2018)
4. Hou, Z.: Research and application of principal component analysis based on interval number theory (2015)
5. Qi, J.F., Xiong, X.P.: Research to the central heating boiler room energy saving index evaluation system. Appl. Mech. Mater. **556–562**, 921–924 (2014)

Selection of Software Development Methodologies (SDMs) Using Bayesian Analysis

Himadri Bhusan Mahapatra, Vishal Chandra and Birendra Goswami

Abstract A software development methodology (SDM)—also called systems development methodology—is a formalized approach for the development of software. Although, there are many different SDMs. In this paper, we proposed a method for the selection of an appropriate SDM for a particular project using Bayesian analysis over various factors affecting the selection of SDM. Bayesian is widely used for decision making. There is a statistical model for various categories established by calculating the conditional probability density functions (PDF). We used some factors, i.e., cost, risk, size, and time duration. We classify categories with the maximum posterior probability within the Bayesian framework. Result of performance of the proposed model is shown by experimentation on the dataset using MATLAB and WINBUGS simulation.

Keywords Bayesian network · SDMs · Probability density function · Frequentist inference · Bayesian inferences

1 Introduction

According to IEEE, a software development methodology is defined as "A framework containing the processes, activities and tasks involved in the development, operation and maintenance of a software product, spanning the life of the system from the definition of its requirements to the termination of its use". It provides a

H. B. Mahapatra (✉)
Department of Computer Science, JRU, Ranchi, India
e-mail: himumahapatra@rediffmail.com

V. Chandra
Department of Computer Science and Engineering, RIT, Koderma, India
e-mail: vcvishalchandra@gmail.com

B. Goswami
Department of Computer Science, ICFAI University, Ranchi, India
e-mail: bgoswami@rediffmail.com

S. Patnaik et al. (eds.), *New Paradigm in Decision Science and Management*, Advances in Intelligent Systems and Computing 1030,
https://doi.org/10.1007/978-981-13-9330-3_2

flexible framework for enhancing the process. It enables effective communication, reuse, and process management [1, 2]. In last 50 years, there are various software development methodologies were developed by various researchers; few of them are: waterfall, iterative waterfall, prototyping model, spiral, evolutionary, incremental, time-boxing, RAD, V-model, etc. Any SDM framework is specific to the type of project we going to develop according to the time duration, complexity, mission-critical problem, resource required, client understandability, and also budget for the project. Therefore, it is the most crucial and most essential task to select the appropriate SDM for the success of project. The success of a software project mainly depends upon the selection of particular software development model for the particular project according to the requirements. Consideration of time and cost is most important when we are going to choose a particular SDM because they play a key role in project development. There are few major parameters which are come to the picture and required for the selection of specific SDM such as Cost, Requirement Specification, Time, Project Size, Project Type, and tolerance to change. The selection of SDM is either based upon the experience or heuristic way, and these approaches sometimes fail due to the involvement of uncertainty.

We have reviewed several literatures which are available on this topic. In [3], authors proposed a rule-based object-oriented framework for SDM selection. Another work proposed in [4], author present a model which was based on knowledge based known as ESPMS for SDM selection. In [5], authors proposed a model known as MODSET which was based on rule-based expert system combined with Likert scale measurement for SDM selection. In this paper, we proposed a method for the selection of an appropriate SDM for a particular project using Bayesian analysis over various factors affecting the selection of SDM. Bayesian analysis is widely used for decision making [6, 7]. There is a statistical model for various categories established by calculating the conditional probability density functions (PDFs). We used some factors' cost, risk, size, and time duration. We classify categories with the maximum posterior probability within the Bayesian framework. Result of performance of the proposed model is shown by experimentation on the dataset.

2 Proposed Work

Our proposed work is based on Bayesian probability distribution. There are various Bayesian inferences [8]; we used a model called Bayesian logistic regression model for the analysis of our dataset. Formula of logistic regression is given below, where **p** is the probability of posterior (likelihood) and **b** is prior distribution, and X_p is parameter.

$$p = \frac{e^{b_0 + b_1 X_1 + b_2 X_2 \ldots + b_p X_p}}{1 + e^{b_0 + b_1 X_1 \ldots + b_p X_p}}$$

MATLAB representation of logistic regression

```
logit = @ (b, x) exp (b (1) +b (2). *x)./(1+exp (b (1) +
b (2). *x))
prior
  prior1 = @(b1) normpdf(b1,0,20;
  likelihood @(b) prod (binopdf (iwf, total projects, logitp
  (b, cost))) … * prior1(b (1)) * prior2(b (2));
```

Propose of this paper is to find the best software development model for a project considering four factors:-

1. Expected cost of project
2. Expected time duration of project
3. Expected risk of failure of project
4. Expected size of project (KLOC).

Expected cost of project:- When a new project starts, first the software development team has to make a rough estimation of the cost of project. This expected cost is determined by the previous experience of similar type of projects. This is estimated in terms of currency.

Expected time duration of project:- Second thing is to estimate how much time is required for completing the whole project. Development team has to determine the time duration to complete the project. This entire thing can be done with the help of the previously completed similar type of project. This factor is estimated in terms of months or weeks.

Expected risk of failure of project:- Every project is associated with a risk of failure of project. Software development team has to estimate the risk of failure of project. This is done by analyzing previously done similar type of projects. This factor is estimated in terms of percentage.

Expected size of project:- After doing all three processes, team decides technology and kilo line of code (KLOC). For example, suppose a new project handover to a team and team decides to do it in a particular language, then the team has to estimate how much lines of code required to do this based on the language. This is also done by using the past experience of similar types of projects. This is estimated in terms of kilo line of code (KLOC).

In this paper, we have compared three software development methodologies (SDMs):-

1. Iterative waterfall model
2. Prototype model
3. V-model.

Data are collected from various companies on software development using three models and four factors.

We draw the conclusion of the selection of SDM with two parameters:

1. Total number of projects
2. One of the four factors.

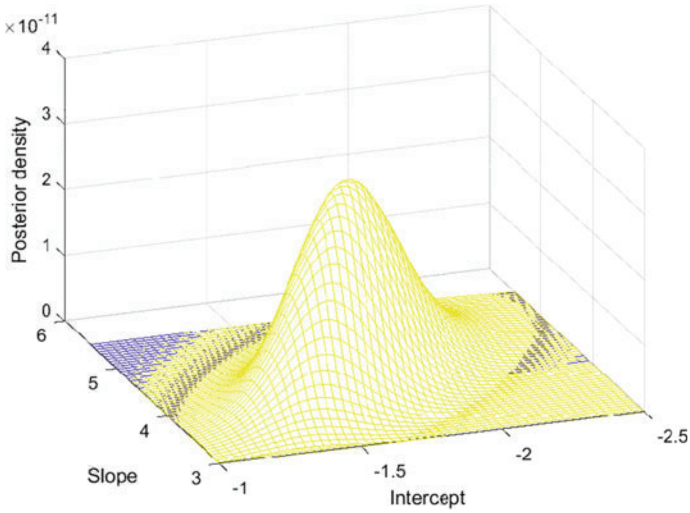

Fig. 1 Surface curve of iterative waterfall, prototype, v-model with cost and total numbers of projects

3 Proposed Model

In this work, we have given an algorithm for the proposed model which is described as follows:

```
model
{
for (i in 1: n) {
# Linear regression on logit
logit(p[i]) <- alpha + b.(factor*factor[i]) +
b. ((no. of projects) *(no. of projects[i]))
# Likelihood function for each data point
frac[i] ~ dbern(p[i])
}
alpha ~ dnorm(0.0,1.0E-4) # Prior for intercept
b. (no. of projects) ~ dnorm (0.0,1.0E-4) # Prior for slope of
number of projects
b. factor ~ dnorm(0.0,1.0E-4) # Prior for slope of factors
}
n=200
```

See Figs. 1, 2, 3 and 4 and Tables 1, 2, 3 and 4.

3.1 Experimental Result

We have collected 200 data from different companies. We used 200 iterations. Results are shown in below table (Tables 5 and 6).

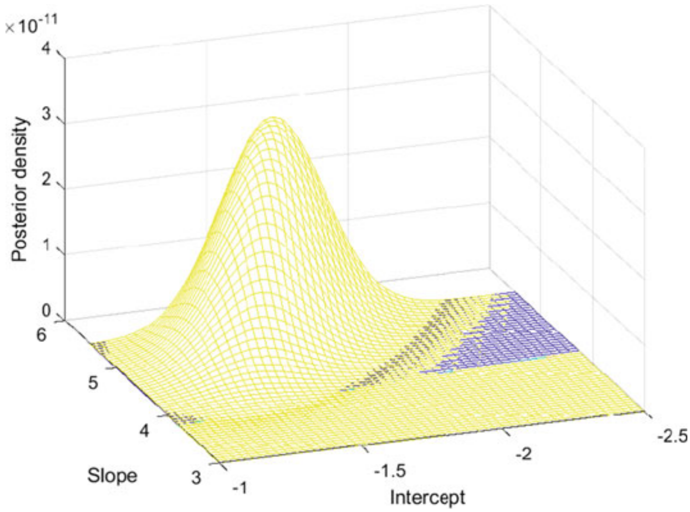

Fig. 2 Shows the surface curve of three SDMs, time, and total number of projects

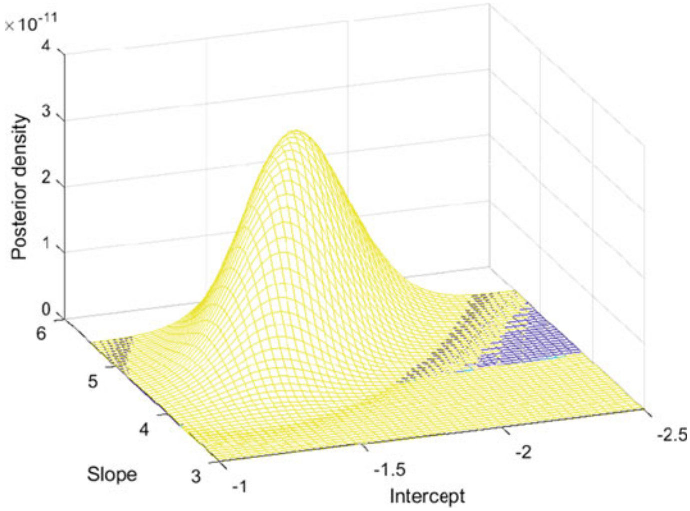

Fig. 3 Shows the surface curve of three SDMs, risk, and total number of projects

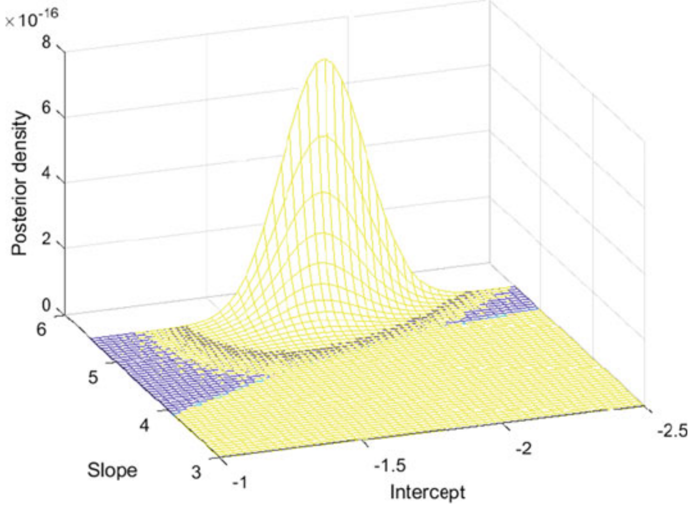

Fig. 4 Shows the surface curve of three SDMs, size, and total number of projects

Table 1 Shows the distribution of cost of projects and total number of projects and their relationship with three SDMs

Cost	Total no. of projects	Prototype model	Iterative waterfall model	V-model
2000	41	1	25	15
2500	42	2	22	18
3000	30	0	19	11
3100	35	6	19	7
5000	31	8	16	7
8000	21	8	9	4

Table 2 Shows the relation between time, total number of projects, and three SDMs

Time (Weeks)	Total no. of projects	Prototype model	Iterative waterfall model	V-model
1	41	0	36	5
3	42	2	20	20
7	30	1	20	9
12	35	6	19	7
15	31	18	5	8
24	21	12	4	5

Table 3 Shows the relation between risk, total number of projects, and three SDMs

Risk in %	Total no. of projects	Prototype model	Iterative waterfall model	V-model
0	41	0	37	4
5	42	2	30	10
8	30	21	4	5
10	35	20	0	15
15	31	24	0	7
20	21	18	0	3

Table 4 Shows the relation between size, total number of projects, and three SDMs

Size in KLOC	Total no. of projects	Prototype model	Iterative waterfall model	V-model
1000	41	0	39	2
1500	42	2	34	6
2300	30	6	16	8
2700	35	10	10	15
4000	31	22	1	8
5000	21	19	0	2

Table 5 Result (proposed model)

No. of projects	Cost	Time	Size	Risk	IWF model (**p**)	Prototype (**p**)	V-model (**p**)
20	2000	2	1000	0	0.91	0.25	0.43
27	1400	3	1400	2	0.78	0.31	0.52
15	4000	4	2000	5	0.64	0.55	0.42
30	3400	3	1600	4	0.79	0.54	0.38
40	4500	5	2300	2	0.49	0.71	0.54
55	6000	10	4000	10	0.26	0.89	0.67
10	1000	2	1200	0	0.79	0.54	0.39

4 Conclusion

On the basis of this research, we can conclude that there is some significant difference between frequentist inference method and Bayesian inference method for the selection of SDM for a particular project. We can also say that Bayesian approach is better than frequentist approach for the selection of SDM. Although proposed model has some degree of uncertainty as in other Bayesian model, it shows better result than other traditional approaches (frequentist). Bayesian method is more flexible and can handle more complex problem which contains uncertainty. The small dataset which