International Journal on Advanced Science, Engineering and Information Technology

#15420 Summary

Submission

Authors	Iwan Tri Riyadi Yanto, Rohmat Saedudin, Sely Novita Sari, Mustafa Mat Deris, Norhalina Senan
Title	Soft Set Multivariate Distribution for Categorical Data Clustering
Original file	15420-34256-1-SM.DOCX 2021-06-02
Supp. files	None
Submitter	Iwan Tri Riyadi Yanto
Date submitted	June 2, 2021 - 10:57 AM
Section	Articles
Editor	Rahmat Hidayat
Abstract Views	9

Status

Status	Published Vol 11, No 5 (2021)
Initiated	2021-10-17
Last modified	2021-10-28

Submission Metadata

Authors

Name	Iwan Tri Riyadi Yanto
Affiliation	Department of Information Systems, University of Ahmad Dahlan, Yogyakarta, Indonesia
Country	Indonesia
Bio Statement	—
Principal contact for editorial correspondence.	
Name	Rohmat Saedudin
Affiliation	Department of Information Systems, Telkom University, Bandung, West Java, Indonesia
Country	Indonesia
Bio Statement	—
Name	Sely Novita Sari
Affiliation	Faculty of Civil Engineering and Planning, Institute Teknologi Nasional Yogyakarta, Indonesia
Country	Indonesia
Bio Statement	—
Name	Mustafa Mat Deris
Affiliation	Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor, Malaysia
Country	Malaysia
Bio Statement	—
Name	Norhalina Senan
Affiliation	Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor, Malaysia
Country	Malaysia
Bio Statement	—

Title and Abstract

Title: Soft Set Multivariate Distribution for Categorical Data Clustering

Abstract

Clustering is the process of breaking down a huge dataset into smaller groups. It has been used in some field studies including pattern recognition, segmentation, and statistics with remarkable success. Clustering is a technique for dividing multivariate datasets into groups. No inherent distance measure on data category makes clustering data more challenging than numerical data. Data category can be assumed following the data from a multinomial distribution. Thus, the standard model parametric model can be used in latent class clustering based on the independent product of multinomial distributions. Meanwhile, multi-valued attributes on the categorical data can be decomposed into the standard set on a multi soft set. In this paper, a clustering technique based on soft set theory is proposed for categorical data through a multinomial distribution. The data will be represented as a multi soft set which is every soft set has its probability of being a member of the cluster. The data with the highest probability will be assigned as the member of the cluster. The experiment of the proposed technique is evaluated based on the Dunn index with regard to the number of clusters and response time. The experiment results show that the proposed technique has the lowest response time with high stability compared to baseline techniques. This study recommends a maximum number of clusters in implementation on the real data.

Indexing

Keywords	Clustering; categorical data; soft set; multivariate.
Language	en

Supporting Agencies

Agencies: —

References

References

C. Wan, M. Ye, C. Yao, and C. Wu, "Brain MR image segmentation based on Gaussian filtering and improved FCM clustering algorithm," in 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017, pp. 1–5.

R. Shanker and M. Bhattacharya, "Brain Tumor Segmentation of Normal and Pathological Tissues Using K-mean Clustering with Fuzzy C-mean Clustering," in VipIMAGE 2017, 2018, pp. 286–296.

A. S. M. S. Hossain, "Customer segmentation using centroid based and density based clustering algorithms," in 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 2017, pp. 1–6.

K. V Ahammed Muneer and K. Paul Joseph, "Performance Analysis of Combined k-mean and Fuzzy-c-mean Segmentation of MR Brain Images," in Computational Vision and Bio Inspired Computing, 2018, pp. 830–836.

H. Zhou, "K-Means Clustering BT - Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods," H. Zhou, Ed. Berkeley, CA: Apress, 2020, pp. 35–47.

S. Irfan, G. Dwivedi, and S. Ghosh, "Optimization of K-means clustering using genetic algorithm," in 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 2017, pp. 156–161.

B. K. D. Prasad, B. Choudhary, and B. Ankayarkanni., "Performance Evaluation Model using Unsupervised K-Means Clustering," in 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 1456–1458.

W. Wei, J. Liang, X. Guo, P. Song, and Y. Sun, "Hierarchical division clustering framework for categorical data," Neurocomputing, vol. 341, pp. 118–134, 2019.

Z. Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values," Data Min. Knowl. Discov., vol. 2, no. 3, pp. 283–304, 1998.

Y. Xiao, C. Huang, J. Huang, I. Kaku, and Y. Xu, "Optimal mathematical programming and variable neighborhood search for k-modes categorical data clustering," Pattern Recognit., vol. 90, pp. 183–195, 2019.

D. B. M. Maciel, G. J. A. Amaral, R. M. C. R. de Souza, and B. A. Pimentel, "Multivariate fuzzy k-modes algorithm," Pattern Anal. Appl., vol. 20, no. 1, pp. 59–71, 2017.

P. S. Bishnu and V. Bhattacherjee, "Software cost estimation based on modified K-Modes clustering Algorithm," Nat. Comput., vol. 15, no. 3, pp. 415–422, 2016.

Z. Huang and M. K. Ng, "A fuzzy k-modes algorithm for clustering categorical data," IEEE Trans. Fuzzy Syst., vol. 7, no. 4, pp. 446–452, 1999.

M. S. Yang, Y. H. Chiang, C. C. Chen, and C. Y. Lai, "A fuzzy k-partitions model for categorical data and its comparison to the GoM model," Fuzzy

Sets Syst., vol. 159, no. 4, pp. 390–405, 2008.

A. Karim, C. Loqman, and J. Boumhidi, "Determining the number of clusters using neural network and max stable set problem," Procedia Comput. Sci., vol. 127, pp. 16–25, 2018.

S. Ben-David, D. Pál, and H. Simon, Stability of k-Means Clustering. 2007.

I. Landi, V. Mandelli, and M. V. Lombardo, "reval: a Python package to determine the best number of clusters with stability-based relative clustering validation," arXiv, vol. 2, no. 4. arXiv, p. 100228, 27-Aug-2020.

D. G. L. Allegretti, "Stability conditions, cluster varieties, and Riemann-Hilbert problems from surfaces," Adv. Math. (N. Y)., vol. 380, p. 107610, Mar. 2021.

E. Andreotti, D. Edelmann, N. Guglielmi, and C. Lubich, "Measuring the stability of spectral clustering," Linear Algebra Appl., vol. 610, pp. 673–697, Feb. 2021.

T. Herawan and M. M. Deris, "On Multi-soft Sets Construction in Information Systems BT - Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence," 2009, pp. 101–110.

D. S. Morris, A. M. Raim, and K. F. Sellers, "A Conway–Maxwell-multinomial distribution for flexible modeling of clustered categorical data," J. Multivar. Anal., vol. 179, p. 104651, 2020.

D.-W. Kim, K. H. Lee, and D. Lee, "Fuzzy clustering of categorical data using fuzzy centroids," Pattern Recognit. Lett., vol. 25, no. 11, pp. 1263–1271, Aug. 2004.

Published by INSIGHT - Indonesian Society for Knowledge and Human Development

UNIVERSITAS
AHMAD DAHLAN

Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

[IJASEIT] Revision Required

1 message

Rahmat Hidayat <rahmat@insightsociety.org> Thu, Jul 22, 2021 at 1:09 PM
To: Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

Iwan Tri Riyadi Yanto:

We have reached a decision regarding your submission to International Journal on Advanced Science, Engineering and Information Technology, "Soft Set Multivariate Distribution for Categorical Data Clustering".

Our decision is to: Revision Required

Editor

Reviewer A:

The manuscript proposes a clustering technique based on soft set theory for categorical data via multinomial distribution. The data are represented as a multi soft set. The experiment of the proposed technique has been evaluated based on the Dunn index to the number of clusters and response time. The experiment results have shown that the proposed technique has a low response time with high stability compared to baseline techniques.

The article title is appropriate and accurately reflects the article's content. The abstract is short, clear, and well-defined. It states the main goal of the paper. The used keywords are appropriate. The introduction is clear, well-written and well-organized. Section 2 presents a study of fuzzy-based clustering theories. It is also referenced with up-to-date literature sources from a suitable range of citations and covering existing relevant works. The used research methodology is well presented in section 3. The obtained results are discussed in detail in section 4. The conclusion is well written and summarize the obtained results of the study.

The manuscript content is structured correctly and contains all the relevant sections marked with subheadings. The figures and tables are enough for the presentation of the data and the results of the research. The manuscript consists of 7 pages, 22 references, 5 sections, 5 figures, and 2 tables. 86% of all cited publications have been published in the last 5 years. The cited literature is from authoritative sources and does not need correction.

Specific comments and suggestions:
• There are some typing errors. The author can reread all the text and correct them. For example, the words: "mutinomial" instead of the "multinomial", "parametic" instead of the "parametric", "clustring" instead of the "clustering", "maksimum" instead of the "maximum", "intance" instead of the "instance" etc.
• There is an empty page at the end of the paper.
• The paper should be formatted according to the IJASEIT template.

International Journal on Advanced Science, Engineering and Information Technology
http://insightsociety.org/ijaseit/index.php/ijaseit

International Journal on Advanced Science, Engineering and Information Technology

#15420 Review

Submission

Authors	Iwan Tri Riyadi Yanto, Rohmat Saedudin, Sely Novita Sari, Mustafa Mat Deris, Norhalina Senan
Title	Soft Set Multivariate Distribution for Categorical Data Clustering
Section	Articles
Editor	Rahmat Hidayat

PeerReview

Round 1

Review Version	15420-34257-1-RV.DOCX 2021-06-02
Initiated	2021-06-06
Last modified	2021-07-02
Uploaded file	None
Editor Version	None
Author Version	15420-34953-1-ED.DOCX 2021-06-30

Round 2

Review Version	15420-34257-2-RV.DOCX 2021-07-02
Initiated	2021-07-02
Last modified	2021-07-14
Uploaded file	None
Editor Version	None
Author Version	15420-34953-2-ED.DOCX 2021-07-23

Round 3

Review Version	15420-34257-3-RV.DOCX 2021-07-27
Initiated	2021-07-27
Last modified	2021-07-31
Uploaded file	None

Editor Decision

Decision	Accept Submission 2021-08-02
Notify Editor	Editor/Author Email Record 2021-08-02
Editor Version	None
Author Version	15420-34953-3-ED.DOCX 2021-09-06 DELETE 15420-34953-4-ED.DOCX 2021-10-27 DELETE
Upload Author Version	Choose File No file chosen Upload

Published by INSIGHT - Indonesian Society for Knowledge and Human Development

UNIVERSITAS
AHMAD DAHLAN

Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

[IJASEIT] Revision Required

1 message

Rahmat Hidayat <rahmat@insightsociety.org> Thu, Jul 22, 2021 at 1:09 PM
To: Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

Iwan Tri Riyadi Yanto:

We have reached a decision regarding your submission to International Journal on Advanced Science, Engineering and Information Technology, "Soft Set Multivariate Distribution for Categorical Data Clustering".

Our decision is to: Revision Required

Editor

Reviewer A:

The manuscript proposes a clustering technique based on soft set theory for categorical data via multinomial distribution. The data are represented as a multi soft set. The experiment of the proposed technique has been evaluated based on the Dunn index to the number of clusters and response time. The experiment results have shown that the proposed technique has a low response time with high stability compared to baseline techniques.

The article title is appropriate and accurately reflects the article's content. The abstract is short, clear, and well-defined. It states the main goal of the paper. The used keywords are appropriate. The introduction is clear, well-written and well-organized. Section 2 presents a study of fuzzy-based clustering theories. It is also referenced with up-to-date literature sources from a suitable range of citations and covering existing relevant works. The used research methodology is well presented in section 3. The obtained results are discussed in detail in section 4. The conclusion is well written and summarize the obtained results of the study.

The manuscript content is structured correctly and contains all the relevant sections marked with subheadings. The figures and tables are enough for the presentation of the data and the results of the research. The manuscript consists of 7 pages, 22 references, 5 sections, 5 figures, and 2 tables. 86% of all cited publications have been published in the last 5 years. The cited literature is from authoritative sources and does not need correction.

Specific comments and suggestions:
• There are some typing errors. The author can reread all the text and correct them. For example, the words: "mutinomial" instead of the "multinomial", "parametic" instead of the "parametric", "clustring" instead of the "clustering", "maksimum" instead of the "maximum", "intance" instead of the "instance" etc.
• There is an empty page at the end of the paper.
• The paper should be formatted according to the IJASEIT template.

International Journal on Advanced Science, Engineering and Information Technology
http://insightsociety.org/ijaseit/index.php/ijaseit

International Journal on Advanced Science, Engineering and Information Technology

#15420 Review

Submission

Authors	Iwan Tri Riyadi Yanto, Rohmat Saedudin, Sely Novita Sari, Mustafa Mat Deris, Norhalina Senan
Title	Soft Set Multivariate Distribution for Categorical Data Clustering
Section	Articles
Editor	Rahmat Hidayat

PeerReview

Round 1

Review Version	15420-34257-1-RV.DOCX 2021-06-02
Initiated	2021-06-06
Last modified	2021-07-02
Uploaded file	None
Editor Version	None
Author Version	15420-34953-1-ED.DOCX 2021-06-30

Round 2

Review Version	15420-34257-2-RV.DOCX 2021-07-02
Initiated	2021-07-02
Last modified	2021-07-14
Uploaded file	None
Editor Version	None
Author Version	15420-34953-2-ED.DOCX 2021-07-23

Round 3

Review Version	15420-34257-3-RV.DOCX 2021-07-27
Initiated	2021-07-27
Last modified	2021-07-31
Uploaded file	None

Editor Decision

Decision	Accept Submission 2021-08-02
Notify Editor	Editor/Author Email Record 2021-08-02
Editor Version	None
Author Version	15420-34953-3-ED.DOCX 2021-09-06 DELETE 15420-34953-4-ED.DOCX 2021-10-27 DELETE
Upload Author Version	Choose File No file chosen Upload

Published by INSIGHT - Indonesian Society for Knowledge and Human Development

UNIVERSITAS
AHMAD DAHLAN

Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

[IJASEIT] Accepted Submission

1 message

Rahmat Hidayat <rahmat@insightsociety.org> Mon, Aug 2, 2021 at 7:06 AM
To: Iwan Tri Riyadi Yanto <yanto.itr@is.uad.ac.id>

Iwan Tri Riyadi Yanto:

We have reached a decision regarding your submission to International Journal on Advanced Science, Engineering and Information Technology, "Soft Set Multivariate Distribution for Categorical Data Clustering".

Our decision is to: Accepted Submission

Rahmat Hidayat
rahmat@insightsociety.org

Reviewer A:

The manuscript proposes a clustering technique based on soft set theory for categorical data via multinomial distribution. The data are represented as a multi soft set. The experiment of the proposed technique has been evaluated based on the Dunn index to the number of clusters and response time. The experiment results have shown that the proposed technique has a low response time with high stability compared to baseline techniques.

The manuscript content is structured correctly and contains all the relevant sections marked with subheadings. The figures and tables are enough for the presentation of the data and the results of the research. The manuscript consists of 7 pages, 22 references, 5 sections, 5 figures, and 2 tables. 86% of all cited publications have been published in the last 5 years.

Specific comments and suggestions:
• I have no suggestions. The paper is well-formatted. The proposed study is presented correctly and contains all needs conclusions.

International Journal on Advanced Science, Engineering and Information Technology
http://insightsociety.org/ijaseit/index.php/ijaseit

International Journal on Advanced Science, Engineering and Information Technology

#15420 Editing

Submission

Authors	Iwan Tri Riyadi Yanto, Rohmat Saedudin, Sely Novita Sari, Mustafa Mat Deris, Norhalina Senan
Title	Soft Set Multivariate Distribution for Categorical Data Clustering
Section	Articles
Editor	Rahmat Hidayat

Copyediting

REVIEW METADATA		REQUEST	UNDERWAY	COMPLETE
1.	Initial Copyedit File: None	—	—	—
2.	Author Copyedit File: None Choose File No file chosen Upload	—	—	
3.	Final Copyedit File: None	—	—	—

Copyedit Comments No Comments COPYEDIT INSTRUCTIONS

Layout

Galley Format		FILE	
1.	PDF VIEW PROOF	15420-37001-1-PB.PDF 2021-10-17	14

Supplementary Files	FILE
	None

Layout Comments No Comments

Proofreading

REVIEW METADATA

		REQUEST	UNDERWAY	COMPLETE
1.	Author	—	—	
2.	Proofreader	—	—	—
3.	Layout Editor	—	—	—

Proofreading Corrections No Comments PROOFING INSTRUCTIONS

Published by INSIGHT - Indonesian Society for Knowledge and Human Development