PSS: New Parametric Based Clustering for Data Category

Iwan Tri Riyadi Yanto[1,3](✉), Mustafa Mat Deris[2], and Norhalina Senan[3]

[1] Department of Information Systems, University of Ahmad Dahlan, Yogyakarta, Indonesia
yanto.itr@is.uad.ac.id

[2] Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
mmustafa@uthm.edu.my

[3] Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
halina@uthm.edu.my

Abstract. This paper proposes a new clustering technique for handling a categorical data called Parametric Soft set (PSS). It bases on statistical distribution namely multinomial multivariate function. The probability of the data category with binary value can be calculated by binomial distribution. Its generalization called multinomial distribution function for data category with multivariate values. Firstly, the data is represented as multi soft set where every object in each soft set has its probability. The probability of each object is calculated by cluster joint distribution function following the multivariate multinomial distribution function. The highest probability will be assigned to the related cluster. The first experiment is conducted to estimate the parameter of the data drawn from random multivariate mixtures distribution. While the second experiment is evaluated the processing times, purity and rand index using benchmarks datasets. The experiment results show that the proposed approach has improved the processing times up to 92.96%. It also has better performance in term of purity and rand index and error mean of the estimation parameters.

Keywords: Clustering · Categorical data · Multi soft set · Multinomial distribution function

1 Introduction

There are two definitions assumed on the partitioning process or clustering process to group the data into several classes. First, well-defined notion of similarity or distance between data objects is needed to measure the resemblance the object. Second, the process to decide the object will be in the same groups or separate into differences group can be developed based on the characteristic of the data [1, 2]. In practice, it called unsupervised learning or clustering process.

R. Ghazali et al. (Eds.): SCDM 2022, LNNS 457, pp. 14–24, 2022.
https://doi.org/10.1007/978-3-031-00828-3_2

There are so many clustering techniques developed because of many various similarity or distance measure in mathematics and many model which can be used to labeling the object such as [3–6]. It makes the notion of clusters cannot be precisely defined and create some various model of clustering i.e. centroid, density, distribution, connectivity, graph-based, neural models, etc. [7]. The clustering technique can be categorized into three types. i.e. pairwise distance cluster, target on optimizing by given merit function and statistical modeling [8]. Only pairwise distances between clustered objects are used in the first type. This is because a tractable mathematical representation for objects is not necessary, these approaches have a wide range of applications. However, due to the quadratic computational complexity of calculating all the pairwise distances, they do not scale well with big data sets. Linkage clustering [9–11] and spectra clustering [12] are two examples. The second type is concerned with optimizing a certain merit function. The merit function represents the widely held idea that good clustering requires objects in the same cluster to be similar, while objects in other clusters should be as diverse as possible. The similarity metric and criterion for evaluating the overall quality of clustering differ amongst algorithms. K-means and k-centroid are two terms that are included in this type. The third type is based on statistical analysis [8]. Each cluster is distinguished by a fundamental parametric distribution (known as a component), such as the multivariate Gaussian for continuous data, the Poisson distribution for discrete data, multinomial distribution for multi values data.

The differences of typical of the data requires careful consideration to determine the similarity or distance measure [2]. In practice, there are various types of data that are used to implement the clustering algorithm, such as numeric, and categorical. Unlike the numerical data, the categorical data contains the attributes which do not have any natural order, so distance measure cannot be executed straightforwardly on categorical attribute [13]. Data category can be assumed following the random multivariate multinomial distribution function [14]. Other hand, categorical data have multi-valued attribute where it can be represented as a multi soft set [15]. Thus, this paper proposes the parametric clustering approach based on soft set theory. The data is decomposed to be a multi soft set respect to all attributes where the probability every soft set in each attribute is calculated using multinomial distribution function. Each object on attributes has different values of probability respect to the cluster. The object with high probability will be assign into the related cluster.

The rest of the paper is organized as follows: Sect. 2 describes related works on information system, soft set, multinomial distribution. Section 3 describes the proposed approach based on soft set multinomial distribution function. Section 4 describes the experiment results on the estimation parameter. Finally, we conclude our work in Sect. 5.

2 Related Works

This section describes the basic of multinomial distribution and soft set theory.

2.1 Multinomial Distribution

Multinomial distribution is a generalization of the binomial distribution [16]. Lets N_i denote the number of results in category i in a sequence of independent trial a with probability p_i for a results in the ith category on each trial, $1 \le i \ge m$, where $\sum_{i=1}^{m} p_i = 1$ Then for every m-tuple of non-negative integers $(n_1, n_2, \ldots, n_m)$ with sum n

$$P(N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m) = \frac{n!}{n_1!n_2!\ldots n_m!} p_1^{n_1} p_2^{n_2} \ldots p_m^{n_m}. \tag{1}$$

A multinomial distribution with parameter $a_k = \left(a_k^{jl}, l = 1, \ldots, m_j, j = 1, \ldots, p\right)$ can be described as the probability mass function as follows;

$$f(x, a_k) = \prod\nolimits_{j=1}^{p} \prod\nolimits_{l=1}^{m_j} \left(a_k^{jl}\right)^{x^{jl}}, \tag{2}$$

where $\sum_{l=1}^{m_j} a_k^{jl} = 1$. The generic polytomous variable $j(j = 1, \ldots, p)$ consist of m_j categories, and $m = \sum_{j=1}^{p} m_j$ indicates the total number of levels.

2.2 Soft Set Theory

Information system can be defined as a tuple $S = (U, A, V, f)$, where U represents the universe of objects, A be a set of variables or parameters, V is a domain (values set) of variable $a \subset A$ where the information function is a total function as in Eq. (3) such that $f(u, a) \in V_a, \forall_{(u,a) \in U \times A}$.

$$f : U \times A \to V. \tag{3}$$

Definition 1. Given $S = (U, A, V, f)$ as an information system. Suppose that $a \in A$, $V_a = \{0,1\}$, then S is a bivalued information system, and can be defined as $S_{\{0,1\}}$.

$$S_{\{0,1\}} = \left(U, A, V_{\{0,1\}}, f\right). \tag{4}$$

Obviously, for every $u \in U, f(u, a) \in \{0, 1\}$, for every $a_i \in A$ and $v \in V$, the map a_i^v of U is $a_i^v : U \to \{0, 1\}$, such that

$$a_i^v = \begin{cases} 1 \, f(u, a) = v \\ 0 \;\; otherwise \end{cases}. \tag{5}$$

An information system can be represented as soft set by handle the uncertainty using an adequate parametrization [17, 18]. Let U be a universe set, E be a set of parameters and $A \subset E$, F is the function that mapping parameter A into the set of all subsets of the set U as in Eq. (6).

$$F : A \to P(U). \tag{6}$$

Then, the pair of (F, A) is called as soft set over $U.\forall_{a\in A}, F(a)$ be considered as the set of a-approximate elements of (F, A).

Consider to an information system definition, a soft set can be interpreted as a special type of information systems, termed a binary-valued information.

Proposition 1. Each Soft set (F, A) can be defined as $\mathrm{S}_{\{0,1\}}$.

Proof: Lets the set of universe U in (F, E) can be considered as the universe U, the set of parameters denoted by E where $A \subset E$. Next, the function of the information system, f is written as:

$$f = \begin{cases} 1, \; u \in F(e) \\ 0, \; u \notin F(e) \end{cases}. \tag{7}$$

That is, when $u_i \in F(e_j)$, where $u_i \in U$ and $e_j \in E$, then $f(u_i, e_j) = 1$, otherwise $f(u_i, e_j) = 0$. To this, we have $V(h_i, e_j) = \{0, 1\}$. Therefore, for $A \subset E$, (F, A) can be represented as $(U, A, V_{\{0,1\}}, f)$. Thus, based on Definition 1, it can be defined as $\mathrm{S}_{\{0,1\}}$.

Definition 2. The value-class of the soft set denoted by $C_{(F,E)}$ are the class of all value sets of a soft set (F, E).

Based on Proposition 1, A Boolean-valued information system deals with the "standard" soft set. For a categorical value of information system denoted by $S = (U, A, V, f)$ with $\cdot V = \bigcup_{a\in A} V_a$ and V_a states the domain of attribute a. The domain V_a has categorical values or multi values. A decomposition can be constructed from S into $|A|$ number of Boolean-valued information system $S = (U, A, V_{\{0,1\}}, f)$. The decomposition of $A = \{a_1, a_2, \cdots, a_{|A|}\}$ into the disjoint-singleton attribute $\{a_1\}, \{a_2\}, \cdots, \{a_{|A|}\}$ is the basis of decomposition of $S = (U, A, V, f)$.

Definition 3. [15] Suppose that $S = (U, A, V, f)$ is a categorical-valued information system and a Boolean-valued information system is expressed by $S = (U, a_i, V_{a_i}, f)$, $i = 1, 2, \cdots, |A|$ with

$$S = (U, A, V, f) = \begin{cases} S^1 = (U, a_1, V_{\{0,1\}}, f) \Leftrightarrow (F, a_1) \\ S^2 = (U, a_2, V_{\{0,1\}}, f) \Leftrightarrow (F, a_2) \\ \quad \vdots \qquad\qquad\qquad\qquad = ((F, a_1), (F, a_2), \cdots, (F, a_{|A|})) \\ S^{|A|} = (U, a_{|A|}, V_{\{0,1\}}, f) \Leftrightarrow (F, a_{|A|}) \end{cases}. \tag{8}$$

Furthermore, a multi soft set over universe U representing a categorical-valued information system $S = (U, A, V, f)$ is expressed as $(F, E) = ((F, a_1), (F, a_2), \cdots, (F, a_{|A|}))$.

3 Proposed Approach

Lets U be an information system random sample following distribution $f(y, \lambda)$. $U = \{u_1, u_2, \ldots, u_{|U|}\}$ will be partition into K cluster $C = \{c_1, c_2, \ldots, c_K\}$ by indicator z_{ik} where $z_{ik} = 1$ if $u_i \in c_k$ and $z_{ik} = 0$ if otherwise. Then, $\prod_{k=1}^{K} \prod_{u_i \in c_k} z_{ik} f_k(y, \lambda)$ is as the cluster joint distribution function of U based on cluster C. For the case data category, the $f(y, \lambda)$ is following the multinomial distribution function. More than that, the categorical data is called a categorical-valued information system $S = (U, A, V, f)$. It can be represented as a multi soft set over of U with $(F, a_1), \cdots, \left(F, a_{|A|}\right) \subseteq (F, A)$ and $\left(F, a_{j_1}, \cdots, \left(F, a_{j|a_j|}\right) \subseteq \left(F, a_j\right)\right)$. Suppose that λ^i_{kjl} is a probability of $u_i \in \left(F, a_{jl}\right)$ into cluster C_k, $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, |U|$, $j = 1, 2, \ldots, |A|$ and $l = 1, 2, \ldots, \left|a_j\right|$. Thus

$$f_k(y, \lambda) = \prod_{j=1}^{|A|} \prod_{l=1}^{|a_j|} \left(\lambda^i_{kjl}\right)^{\left|F, a_{j_l}\right|}, \textit{where} \sum_{l=1}^{|a_j|} \lambda_{kjl} = 1, \forall k, j \tag{9}$$

By substituting the multinomial distribution function into the cluster joint distribution function, then the maximum objective function is defined as

$$\begin{aligned} \textit{Maximize}L_{CML}(z, \lambda) &= \sum_{k=1}^{K} \sum_{i=1}^{|U|} z_{ik} \prod_{j=1}^{|A|} \prod_{l=1}^{|a_j|} \left(\lambda^i_{kjl}\right)^{\left|F, a_{j_l}\right|} \\ &= \sum_{k=1}^{K} \sum_{i=1}^{|U|} z_{ik} \sum_{j=1}^{|A|} \sum_{l=1}^{|a_j|} \ln\left(\lambda^i_{kjl}\right)^{\left|F, a_{j_l}\right|}. \end{aligned} \tag{10}$$

Subject to

$$\sum_{k=1}^{K} z_{ik} = 1, \mathrm{z_{ik}} \in \{0, 1\} \textit{for } i = 1, 2, \ldots, |U| \textit{ and } \sum_{l=1}^{|a_j|} \lambda_{kjl} = 1.$$

The solution of the objective function are

$$\lambda_{kjl} = \frac{\sum_{u_i \in \left(F, a_{j_l}\right)} z_{ik}}{\sum_{l=1}^{|a_j|} \sum_{u_i \in \left(F, a_{j_l}\right)} z_{ik}}, \tag{11}$$

$$z_{ik} = \begin{cases} 1 \textit{ if } \sum\limits_{j=1}^{|A|} \ln\left(\lambda^i_{kjl}\right) = \max\limits_{1 \le k' \le K} \sum\limits_{j=1}^{|A|} \ln\left(\lambda^i_{kjl}\right). \\ 0 \qquad \textit{otherwise} \end{cases} \tag{12}$$

4 Experimental Results

In this section, the computational complexity is analyzed to show the computational cost of the proposed approach compared to the baseline technique. The experiment is conducted to estimate the parameter of the data drawn from random multivariate mixtures distribution and evaluated the processing times, purity and rand index using benchmarks datasets.

4.1 Computational Complexity

The PSS technique need $O(J)$ to decompose multi soft set (F, A), and parameter λ^i_{kjl} and indicator z_{ik} are $O(KM)$ and $O(KIM)$, respectively. Therefore, the polynomial of $O(KM(I+1)t+J)$ is a form of the computational complexity for the PSS technique. For fuzzy centroid technique, the computational cost for update fuzzy centroid v_{kjl} is $O(KIM)$ and partition matrix μ_{ik} is $O(KIJ)$, in each iteration. Thus, the complexity of fuzzy centroid is $O(KI(M+J)t)$. Meanwhile, fuzzy k-partition need $O(KIM)$ and $O(KIM)$ to update the parameters λ_{kjl} and fuzzy partition μ_{ik}, respectively, thus the overall is $O(2KIMt)$. Thus, the PSS technique has the smallest computational complexity, and it can be seen as in Table 1.

Table 1. A comparison results of three algorithms in computational complexity

Algorithms	Computational complexity
FC [4]	$O(2KIMt)$
FkP [19]	$O(2KIMt)$
PSS	$O(KM(I+1)t+J)$

4.2 Estimation Parameters

In this part, the experiments are designed to estimate the parameters of multivariate multinomial mixtures. The data samples are assumed and generated randomly using distribution function $f(y, \lambda)$

$$f(y,\lambda) = \sum\nolimits_{k=1}^{K} \alpha_k f_k(y,\lambda)\, with f_k(y,\lambda) = \prod\nolimits_{j=1}^{J} \prod\nolimits_{l=1}^{L_j} \left(\lambda_{kjl}\right)^{y_{jl}}. \tag{13}$$

It is called as the mixture distribution. The mixing proposition α_k are estimated by $\alpha_k = \sum_{i=1}^{I} \frac{g_{ik}}{I}, \alpha_k = \sum_{i=1}^{I} \frac{\mu_{ik}}{I}, \alpha_k = \sum_{i=1}^{I} \frac{z_{ik}}{I}, k = 1, \ldots, K$, where g_{ik}, μ_{ik} and z_{ik} are the results of the Fuzzy Centroid, Fuzzy k-partition and the PSS approach, respectively.

In this experiment, the data are taken from the multinomial binomial mixture distribution as in [19]. The algorithms are used to estimate the parameters of a four-variable binomial mixture of two classes as in Eq. (13). Thus, the data are create randomly from the mixture distribution $f(y, \lambda)$ with

$$\begin{aligned} f(y,\lambda) = \sum_{k=1}^{2} \alpha_k f_k(y,\lambda_k) &= \alpha B(1,\lambda_{11}) B(1,\lambda_{12}) B(1,\lambda_{13}) B(1,\lambda_{14}) \\ &+ (1-\alpha) B(1,\lambda_{21}) B(1,\lambda_{22}) B(1,\lambda_{23}) B(1,\lambda_{24}), \end{aligned} \tag{14}$$

where $B(1, p)$ is a Bernoulli distribution with four different α and two different λ.

$$\alpha_1 = (0.30.7), \alpha_2 = (0.40.6), \alpha_3 = (0.10.9), \alpha_4 = (0.70.3), \alpha_4 = (0.90.1)$$

$$\lambda_a = \begin{pmatrix} 0.7\ 0.5\ 0.8\ 0.6 \\ 0.9\ 0.3\ 0.4\ 0.4 \end{pmatrix}, \quad \lambda_b = \begin{pmatrix} 0.3\ 0.4\ 0.5\ 0.6 \\ 0.6\ 0.5\ 0.4\ 0.3 \end{pmatrix}$$

There are 200 random samples for all combinations from mixture distribution and random initial values used in the implementation of this algorithm. Each technique is run 10 times. In this experiment, MATLAB version 9.0.0.341360 (R2016a) is used to determine the performance in terms of time responses (tr) and Mean Square Error (MSE) of parameters of both the estimates and true parameters. They are executed sequentially on the specifications of a computer with an Intel Core i5, the total main memory is 8 GB, and the operating system is Mac OS High Sierra. Table 2 shows the achievement of mean square error and response time of the techniques. The average of mean square of the PSS are 0.0848 and 0.0867 in term of estimating parameter λ_a in all a combinations. It is lowest than the fuzzy centroid and fuzzy K partition. Meanwhile for the estimate parameter λ_a in all a combinations, the PSS get 0.0813 and 0.0838, a little higher than the fuzzy K partition and fuzzy centroids. However, in average the PSS able to improve up to 5.92%, 2.25% in term of estimating λ and α, respectively, and up to 92.96% in term of response times as in summarized in Table 3. These results show that PSS possesses more efficient and accurate.

Table 2. Average response times and MSE for all experiments

	Fuzzy centroid			Fuzzy K partition			PSS		
λ_a	λ	α	tr	λ	α	tr	λ	α	tr
a_1	0.0840	0.0502	0.0808	0.0821	0.0612	0.0923	0.0834	0.0482	0.0061
a_2	0.1096	0.0136	0.0982	0.1024	0.0127	0.0872	0.1046	0.0155	0.0062
a_3	0.0419	0.1979	0.0512	0.0641	0.1800	0.0562	0.0499	0.1557	0.0042
a_4	0.1101	0.0394	0.1019	0.1047	0.0310	0.1086	0.1053	0.0304	0.0091
a_5	0.1058	0.1792	0.0827	0.0886	0.1900	0.0608	0.0808	0.1836	0.0040
Average	0.0903	0.0961	0.0884	0.0884	0.0950	0.0810	**0.0848**	**0.0867**	**0.0059**
λ_b	λ	α	tr	λ	α	tr	λ	α	tr
a_1	0.0864	0.0230	0.0687	0.0837	0.0292	0.0757	0.0779	0.0408	0.0052
a_2	0.1111	0.0102	0.0846	0.1092	0.0148	0.0879	0.1165	0.0153	0.0067
a_3	0.0597	0.1697	0.0604	0.0877	0.1584	0.0554	0.0704	0.1690	0.0037
a_4	0.1044	0.0411	0.0637	0.0900	0.0510	0.0816	0.0741	0.0447	0.0048
a_5	0.1075	0.1737	0.0453	0.0704	0.1436	0.0392	0.0677	0.1491	0.0027
Average	0.0938	0.0835	0.0645	0.0882	0.0794	0.0680	0.0813	0.0838	0.0046

Table 3. The improvement of the results

Parameter	Fuzzy centroid	Fuzzy K partition	PSS	Improvement
λ	0.0920	0.0883	0.0831	5.92%
α	0.0898	0.0872	0.0852	2.25%
tr	0.0738	0.0745	0.0053	92.96%

4.3 Applying to the Dataset

Three categorical datasets obtained from the UCI Machine Learning Repository [20], namely Soybean, Tic-tac-toe, and Balloons are used. Table 4, describes the dataset used which consists of dataset number, the name of the dataset, the number of attributes and the number of dataset instances. The 100 of FC, FkP and PSS implementation for all dataset are run to give different random initial membership function. The average in term of cluster purity and Rank Index are captured in Fig. 1 and Fig. 2, respectively. It shows that the PSS have better performance comparing to FC and FkP techniques on soybean dataset. Even, on the balloon and tic-tac-toe datasets the PSS has close similar result. Is shows that the PSS technique is able to maintain the cluster purity and Rank index compared by the FC and FkP. Nevertheless, The result of time response as shown in Fig. 3 indicates that PSS overcome FC and FkP technique. In detail, FC and FkP respectively consume approximately 0.6161 s and 0.7062 s of execution time to Process three datasets in average. In contrast, PSS technique requires only approximately 0.0193 s of execution time in average for three datasets. It clearly shows a reduction of execution time by 0.6419 s in average. Thus the PSS is superior in terms of computational time with able to maintenance the rank index and purity comparing to the baselines.

Table 4. Dataset used for experiments

No	Dataset name	#Attributes	#Instances
1	Soybean	35	47
2	Balloon	4	20
3	Tic-tac-toe	9	958

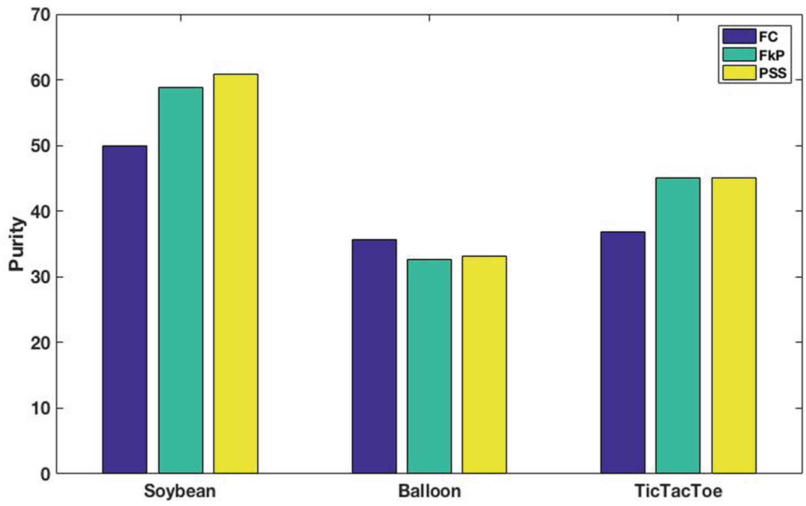

Fig. 1. Experimental result of the cluster purity

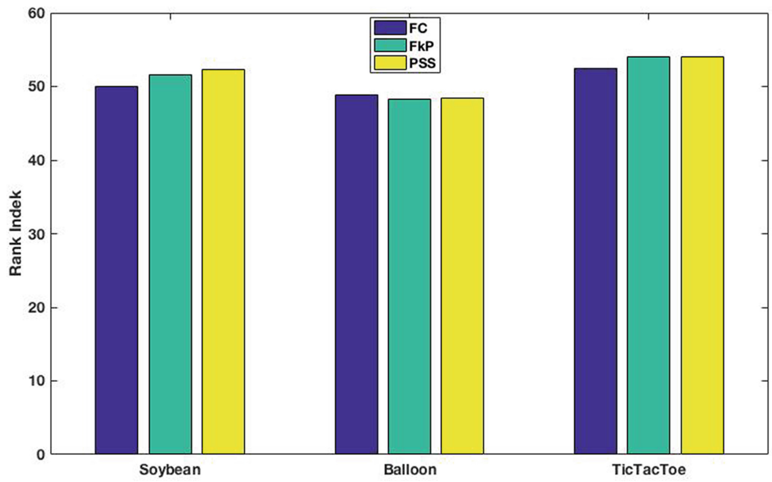

Fig. 2. Experimental result of the rand index

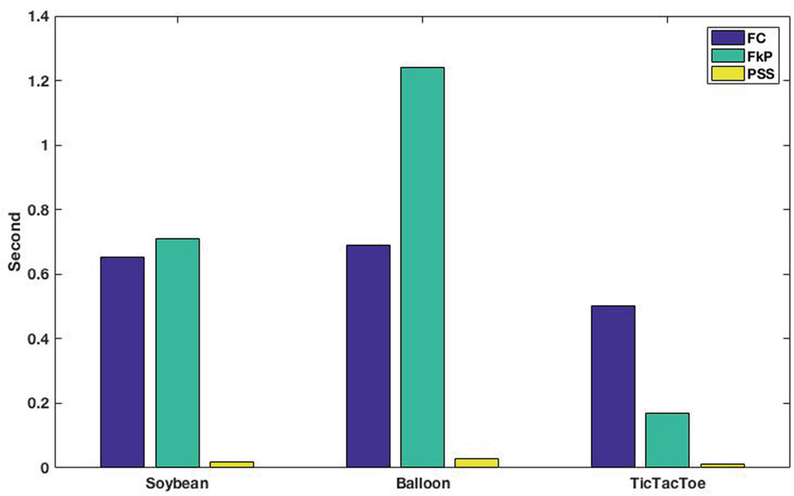

Fig. 3. Experimental result of time responses

5 Conclusion

The new parametric clustering technique has been proposed for data category. It is applied cluster joint distribution assumed following the multinomial multivariate distribution. The highest probability will be assigned to the related cluster. Comparative analysis of the proposed algorithm called PSS and two baseline algorithms with respect to error mean of the statistically testing is carried out to estimate the parameter and response times. The results show that the proposed approach outperforms the existing approaches in terms of lower response times up 92.96%. Then, the techniques is implemented to benchmark datasets to know the performance in terms of Rank index, Purity and time responses. The experiment shows that the PSS technique is out performance compared to the baseline techniques.

References

1. Haixia, X.U., Zheng, T.: An optimal spectral clustering approach based on Cauchy-Schwarz divergence. Chin. J. Electron. **18**(1), 105–108 (2009)
2. Von Luxburg, U., Williamson, R.C., Guyon, I.: Clustering: science or art? J. Mach. Learn. Res. Proc. Track **27**, 65–80 (2012)
3. Leopold, N., Rose, O.: UNIC: a fast nonparametric clustering. Pattern Recognit. **100**, 107117 (2020)
4. Nooraeni, R., Arsa, M.I., Kusumo Projo, N.W.: Fuzzy centroid and genetic algorithms: solutions for numeric and categorical mixed data clustering. Procedia Comput. Sci. **179**(2020), 677–684 (2021)
5. Golzari Oskouei, A., Balafar, M.A., Motamed, C.: FKMAWCW: categorical fuzzy k-modes clustering with automated attribute-weight and cluster-weight learning. Chaos Solitons Fractals **153**, 111494 (2021)
6. Kuo, R.J., Zheng, Y.R., Nguyen, T.P.Q.: Metaheuristic-based possibilistic fuzzy k-modes algorithms for categorical data clustering. Inf. Sci. (Ny) **557**, 1–15 (2021)
7. Hennig, C.: What are the true clusters? Pattern Recognit. Lett. **64**, 53–62 (2015)
8. Li, J., Ray, S., Lindsay, B.G.: A nonparametric statistical approach to clustering via mode identification. J. Mach. Learn. Res. **8**, 1687–1723 (2007)
9. Ramos Emmendorfer, L., de Paula Canuto, A.M.: A generalized average linkage criterion for hierarchical agglomerative clustering. Appl. Soft Comput. **100**, 106990 (2021)
10. Bi, X., Luo, X., Sun, Q.: Branch tire packet classification algorithm based on single-linkage clustering. Math. Comput. Simul. **155**, 78–91 (2019)
11. Schmidt, M., Kutzner, A., Heese, K.: A novel specialized single-linkage clustering algorithm for taxonomically ordered data. J. Theor. Biol. **427**, 1–7 (2017)
12. Xiong, Y., et al.: A spectra partition algorithm based on spectral clustering for interval variable selection. Infrared Phys. Technol. **105**, 103259 (2020)
13. Nguyen, T.P.Q., Kuo, R.J.: Partition-and-merge based fuzzy genetic clustering algorithm for categorical data. Appl. Soft Comput. J. **75**, 254–264 (2019)
14. Sinharay, S.: Discrete probability distributions. Int. Encycl. Educ., 132–134 (2010). https://doi.org/10.1016/B978-0-08-044894-7.01721-8
15. Herawan, T., Deris, M.M.: On multi-soft sets construction in information systems. In: Huang, DS., Jo, KH., Lee, HH., Kang, HJ., Bevilacqua, V. (eds.) ICIC 2009. LNCS, vol. 5755. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04020-7_12
16. Malefaki, S., Iliopoulos, G.: Simulating from a multinomial distribution with large number of categories. Comput. Stat. Data Anal. **51**(12), 5471–5476 (2007)

17. Molodtsov, D.: Soft set theory—first results. Comput. Math. with Appl. **37**(4–5), 19–31 (1999)
18. Maji, P.K., Biswas, R., Roy, A.R.: Soft set theory. Comput. Math. with Appl. **45**(4–5), 555–562 (2003)
19. Yang, M.S., Chiang, Y.H., Chen, C.C., Lai, C.Y.: A fuzzy k-partitions model for categorical data and its comparison to the GoM model. Fuzzy Sets Syst. **159**(4), 390–405 (2008)
20. Dheeru, D., Karra Taniskidou, E.: UCI Machine Learning Repository (2017)