

Article

A Possible World-Based Fusion Estimation Model for Uncertain Data Clustering in WBNs

Chao Li ¹, Zhenjiang Zhang ^{2,*}, Wei Wei ³, Han-Chieh Chao ⁴ and Xuejun Liu ⁵

¹ Department of Electronic and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China; 15111037@bjtu.edu.cn

² The School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

³ Shaanxi Key Laboratory for Network Computing and Security Technology, School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; weiwei@xaut.edu.cn

⁴ Department of Electrical Engineering, National Dong Hwa University, Hualien 97401, Taiwan; hcc@mail.ndhu.edu.tw

⁵ School of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China; lxj@bipt.edu.cn

* Correspondence: zhangzhenjiang@bjtu.edu.cn

Abstract: In data clustering, the measured data are usually regarded as uncertain data. As a probability-based clustering technique, possible world can easily cluster the uncertain data. However, the method of possible world needs to satisfy two conditions: determine the data of different possible worlds and determine the corresponding probability of occurrence. The existing methods mostly make multiple measurements and treat each measurement as deterministic data of a possible world. In this paper, a possible world-based fusion estimation model is proposed, which changes the deterministic data into probability distribution according to the estimation algorithm, and the corresponding probability can be confirmed naturally. Further, in the clustering stage, the Kullback–Leibler divergence is introduced to describe the relationships of probability distributions among different possible worlds. Then, an application in wearable body networks (WBNs) is given, and some interesting conclusions are shown. Finally, simulations show better performance when the relationships between features in measured data are more complex.

Keywords: possible worlds; fusion estimation; uncertain data; clustering



Citation: Li, C.; Zhang, Z.; Wei, W.; Chao, H.-C.; Liu, X. A Possible World-Based Fusion Estimation Model for Uncertain Data Clustering in WBNs. *Sensors* **2021**, *21*, 875. <https://doi.org/10.3390/s21030875>

Academic Editor: Dawid Połap

Received: 15 December 2020

Accepted: 25 January 2021

Published: 28 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Clustering is a kind of technology for machine learning that puts similar objects into the same cluster. Clustering techniques play an important role in many areas such as health care and action recognition in the medical domain [1,2], behavior surveillance and battlefield prediction in the military field [3,4], resource and information management in the communications field [5,6], and so on. There are plenty of cluster methods presented that can be divided into three principal types according to the clustering scale: distance-based, density-based, and connectivity-based [7,8].

Most clustering methods focus on deterministic data. Unfortunately, almost all clustering data are collected by the corresponding equipment, which entails measuring errors. In this case, the uncertain data can describe the measurement data better. For acquiring better and more appropriate results, the fusion estimation methods such as the Bayes-based [9], Kalman-based [10], or artificial intelligence-based [11,12] methods are commonly used to estimate the measurements.

Fusion estimation is a technology that uses the computing power of data acquisition equipment to de-noise and de-redundancy the measurement data according to certain rules. It focuses on mining data information, designing corresponding estimation algorithms,

and improving the accuracy of data. In this technology, the measurement data are de-noised first, and then the data is fused on the time series to obtain the accurate conclusion for uncertain data. Finally, the uncertain data is processed by clustering, and the final processing result is obtained.

Many methods have been proposed to deal with uncertain data in recently years [13–15]. Among these methods, the possible world-based methods have been demonstrated to be efficient and reasonable. Possible world-based clustering methods consider all the probabilities of the uncertain data and fuse them into the final clustering result. This kind of method usually exhibits good performance. On the other hand, the uncertain data can be represented by a probability distribution in most cases. Therefore, the Kullback–Leibler divergence (KL divergence) [16] is used to describe the similarity of two probability distributions.

In practice, there are differences in the accuracy of different acquisition equipment, which is represented by the differences in data uncertainty. Existing algorithms based on possible worlds can deal with the difference problem of uncertainty in a relatively simple way. In this paper, variance, an important statistic of data uncertainty, is introduced into the model of possible worlds to study its role in improving accuracy. Then, a possible world-based fusion estimation model (PWFEM) for uncertain data is presented, which includes two methods according to different distance-based formulas. When the variance of uncertain data is small, the numerical distance-based method (PWFEM-nd) is employed. The probabilistic distance-based method (PWFEM-pd) is employed when variance is prominent. Then, the application in wearable body networks (WBNs) is introduced. The specific derivation formula is given with the different distance-based formulas. Finally, the simulations show good performance in terms of the proposed model.

The rest of the paper is organized as follows. In Section 2, the related works are introduced. In Section 3, the preliminaries are introduced, and some definitions and assumptions are given. The theoretical derivation of the PWFEM is given in Section 4. In Section 5, the simulations examine the performance of the PWFEM. Finally, conclusions are given in Section 6.

2. Related Works

In this section, the processing technologies of uncertain data are introduced in detail. The collected data that come from acquisition equipment contain noise, which means the collected data contain great uncertainty. Therefore, it is necessary to perform fusion estimation processing on the data first, and use the rules and redundancy of the data itself to improve the data accuracy and reduce the uncertainty of the data.

Commonly used fusion estimation algorithms include Bayes filter (BF) [17], Kalman filter (KF) [18], extended Kalman filter (EKF) [19], unscented Kalman filter (UKF) [20], and particle filter (PF) [21]. Wherein, BF and KF are estimates of linear systems, BF can theoretically estimate data of arbitrary noise distribution, and KF is BF when the noise is Gaussian white noise. The EKF, UKF, and PF are the estimates of the nonlinear system, where EKF is for weak nonlinear system, UKF is for strong nonlinear systems and has high computing complexity, while the PF is calculated directly from the average probability density conditions, in which the probability density is determined by EKF and UKF approximation, but the estimation precision is higher than that of a single use of EKF or UKF, but the number of calculations is much higher than that of EKF and UKF.

In [22], the authors argued that two possible world-based clustering algorithms suffered from the following issues: (1) they dealt with each possible world independently and ignored the consistency principle across different possible worlds; (2) they required an extra post-processing procedure to obtain the results, which meant that effectiveness was highly dependent on the post-processing method, and their efficiencies were also not very good. In order to solve the problems above, Liu et al. proposed a possible world-based consistency learning model that considered the consistency principle during the clustering/classification procedure and thus could achieve satisfactory performance.

The Possible world based consistency learning model for clustering uncertain data (PWCLU) was proposed in [22], which holds that the clustering results in each possible world are consistent. Several equipment types were used for collecting the same data. Each piece of data for one piece of equipment was considered to belong to a possible world, and the probability was regarded as equal for each possible world. The authors only gave an algorithm to deal with finite possible worlds.

On the other hand, clustering algorithms usually require a method to describe the distance between two datasets. In uncertain data, the distance can be expressed as a probability distribution in most cases. Therefore, a method of describing the distance between probability distributions is required. Sinkkonen and Kaski [23] studied the problem of learning groups or categories that were local in the continuous primary space but homogeneous according to the distributions of an associated auxiliary random variable over a discrete auxiliary space. In their model, Kullback–Leibler divergence was used to calculate the distance between two probability distributions.

In this paper, a possible world-based fusion estimation model (PWFEM) is proposed for clustering uncertain data. The proposed model removes the assumption of the consistency principle of [22]. Moreover, two PWFEM-based methods are given. One generalizes the PWCLU to the continuous possible worlds, which is based on numerical distance. Therefore, it is called PWFEM-nd. The other is based on probability distribution distance and is named PWFEM-pd. Then, an application in WSNs is discussed. Two specific distance functions that correspond to the numerical distance and probability distribution distance, respectively, are introduced to prove that the PWFEM-nd is equivalent to PWFEM-pd under certain circumstances. Finally, the simulations are discussed; they showed good performance of the models.

3. Preliminaries

In this section, some necessary definitions and assumptions are given for possible world and Kullback–Leibler divergence; the assumptions of independence for each component of the datasets and the structure of the data are also given.

3.1. Definition of Possible World

Let $O \in R^{N \times n}$, $O = \{O_1, O_2, \dots, O_n\}$ be an uncertain dataset, where O is not deterministic data but a probability distribution. If O is a discrete probability distribution, pw is one of the possibilities of the uncertain data O , which can be written as $pw = \{O_1^{pw}, O_2^{pw}, \dots, O_n^{pw}\}$, which is deterministic data with its probability $P(pw)$. If O is a continuous probability distribution, O can be described as a probability density function $f(pw)$, where pw is the value of the random variable O . Then,

$$\int_{\mathcal{D}} f(pw) dpw = 1$$

3.2. Definition of Kullback–Leibler Divergence

Let $p(x)$ and $q(x)$ be the distribution of random variable X , so the Kullback–Leibler divergence of $p(x)$ and $q(x)$ is:

$$d_{KL}(p(x), q(x)) = \int_{-\infty}^{+\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \quad (1)$$

3.3. Some Assumptions

Assumption 1. Almost all possible worlds exhibit the same class labels and cluster structures, and they exhibit the different class labels and cluster structures with small probabilities.

Assumption 2. In Section 5, it is assumed that $\forall x_i, x_j \in X, x_i + x_j$ is also the Gaussian distribution.

Assumption 3. In Section 5, it is assumed that the wearable nodes keep a stable state to collect the data all the time. Therefore, the covariance matrix will not change.

4. Possible World-Based Fusion Estimation Model (PWFWM)

In this section, the details of the PWFWM are introduced in three parts. The first part is the introduction of data fusion estimation. The second part is the introduction of the calculation process of distribution distance. The third part introduces the clustering method based on the possible world.

4.1. Data Fusion Estimation

The collected data can be divided into two types: filterable data and high accuracy data. Without loss of generality, it is assumed the measurement data at time t is:

$$M_t = [z_1^f, z_2^f, \dots, z_q^f, z_1^a, z_2^a, \dots, z_s^a]_t \quad (2)$$

where $M_t^f = [z_1^f, z_2^f, \dots, z_q^f]_t$ are the filterable data, and $M_t^a = [z_1^a, z_2^a, \dots, z_s^a]_t$ are the high-accuracy data.

Corresponding to the possible world, filterable data are the probabilistic data, while the high accuracy data are the numeric data. It is assumed the format of the clustering data in a possible world at time t is:

$$X_t = [x_1^p, x_2^p, \dots, x_h^p, x_1^n, x_2^n, \dots, x_s^n]_t \quad (3)$$

where $X_t^p = [x_1^p, x_2^p, \dots, x_h^p]_t$ are the probability data, and $X_t^n = [x_1^n, x_2^n, \dots, x_s^n]_t$ are the numeric data.

In most cases, the filterable data can be obtained according to the Kalman-based filter. The high accuracy data can be converted to filterable data by the Gaussian distribution, whose expectation is zero and whose variance is small. The details are as follows.

The measurement data are first converted to the clustering data by the following formulas:

If the filterable data satisfied the following state function and measurement function:

$$\begin{cases} X_t^p = f(X_{t-1}^p) + \omega_{t/t-1} \\ M_t^f = g(X_{t-1}^p) + v_t \end{cases} \quad (4)$$

The appropriate filter algorithm can be used to solve the functions above. If the result is \hat{X}_t^p , the probability data can be written as $\hat{X}_t^p + \omega_{t/t-1}$.

Similarly, the numerical data can be written as $X_t^n = M_t^a + \omega_t^a$, where ω_t^a is Gaussian distribution with zero mean and small variance.

Then, we have:

$$X_t = \begin{bmatrix} \hat{X}_t^p + \omega_{t/t-1} \\ M_t^a + \omega_t^a \end{bmatrix} = \begin{bmatrix} \hat{X}_t^p \\ M_t^a \end{bmatrix} + \begin{bmatrix} \omega_{t/t-1} \\ \omega_t^a \end{bmatrix} = \hat{X}_t + \Omega_t \quad (5)$$

where $\hat{X}_t = \begin{bmatrix} \hat{X}_t^p \\ M_t^a \end{bmatrix}$ and $\Omega_t = \begin{bmatrix} \omega_{t/t-1} \\ \omega_t^a \end{bmatrix}$. Moreover, we let $\Omega_t = \begin{bmatrix} \omega^p \\ \omega^a \end{bmatrix}$, which is a scleronomic Gaussian distribution. Therefore, according to Assumption 2, the multivariate Gaussian distribution with X_t can be written as follows:

$$X = \frac{1}{(\sqrt{2\pi})^l |\Sigma|^{\frac{1}{2}}} e^{-\frac{(x-\mu_x)^T (\Sigma)^{-1} (x-\mu_x)}{2}} \quad (6)$$

where $l = h + s$, $\mu_x = [E(x_i)]_{i=1}^l$ and

$$\Sigma = [\sigma_{ij}]_{l \times l}, \sigma_{ij} = \sqrt{D(x_i) \cdot D(x_j)} \quad (7)$$

Based on the above, the structure of clustering data can be confirmed. Then, the distance-based functions need to be confirmed.

4.2. Distance Calculation Method Based on KL Divergence-Based Distance

Almost all clustering algorithms need to calculate the distance. In the PWFWM, there are two types of data: filterable and high accuracy. For accuracy data, the Euclidean distance can be used, and the KL divergence can be used to process the filterable data. In this Section, the distance calculation method based on KL divergence is introduced in detail.

KL divergence analyzes the degree of difference between two distributions from the perspective of information entropy. Assume that $p(x)$ and $q(x)$ are two distributions of random variable X , then the KL divergence is:

$$KL(p \parallel q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (8)$$

The calculation formula in the discrete case is:

$$KL(p \parallel q) = \sum_{i=1}^n p(x_i) \log \frac{p(x_i)}{q(x_i)} \quad (9)$$

Assuming that the probability distribution is usually Gaussian, $P \sim N(\mu_1, \Sigma_1)$ and $Q \sim N(\mu_2, \Sigma_2)$, and the dimension of the data is n . Then, the KL divergence calculation formula is as follows:

$$KL(P \parallel Q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx = E_P[\log p(x) - \log q(x)] \quad (10)$$

Plugs the $P \sim N(\mu_1, \Sigma_1)$ and $Q \sim N(\mu_2, \Sigma_2)$ in (10):

$$\begin{aligned} KL(P \parallel Q) &= \frac{1}{2} E_P \left[\log \frac{|\Sigma_2|}{|\Sigma_1|} - (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) + (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right] \\ &= \frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2} E_P \left[(x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right] + \frac{1}{2} E_P \left[(x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right] \end{aligned} \quad (11)$$

where

$$\begin{aligned} E_P \left[(x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right] &= E_P \left[\text{tr} \left(\Sigma_1^{-1} (x - \mu_1) (x - \mu_1)^T \right) \right] \\ &= \text{tr} \left[E_P \left(\Sigma_1^{-1} (x - \mu_1) (x - \mu_1)^T \right) \right] = \text{tr} \left[\Sigma_1^{-1} E_P \left((x - \mu_1) (x - \mu_1)^T \right) \right] = n \end{aligned} \quad (12)$$

and

$$\begin{aligned} E_P \left[(x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right] &= E_P \left[\text{tr} \left(\Sigma_2^{-1} (x - \mu_2) (x - \mu_2)^T \right) \right] \\ &= \text{tr} \left[\Sigma_2^{-1} E_P \left((x - \mu_2) (x - \mu_2)^T \right) \right] = \text{tr} \left[\Sigma_2^{-1} E_P (xx^T - x\mu_2^T - \mu_2 x^T + \mu_2 \mu_2^T) \right] \\ &= \text{tr} \left[\Sigma_2^{-1} (\Sigma_1 + \mu_1 \mu_1^T - \mu_1 \mu_2^T - \mu_2 \mu_1^T + \mu_2 \mu_2^T) \right] = \text{tr} \left[\Sigma_2^{-1} \Sigma_1 + \Sigma_2^{-1} (\mu_1 - \mu_2) (\mu_1 - \mu_2)^T \right] \\ &= \text{tr} \left(\Sigma_2^{-1} \Sigma_1 \right) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) \end{aligned} \quad (13)$$

Finally, we have

$$KL(P \parallel Q) = \frac{1}{2} \left[\log \frac{|\Sigma_2|}{|\Sigma_1|} - n + \text{tr} \left(\Sigma_2^{-1} \Sigma_1 \right) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) \right] \quad (14)$$

Moreover, if $\Sigma_1 = \Sigma_2 = \Sigma$. Then, we get:

$$d_{KL}(i, j) = KL(P \parallel Q) = \frac{1}{2}(u_j - u_i)^T \Sigma^{-1} (u_j - u_i) \quad (15)$$

In this way, the distance between two probability distributions is obtained. Then, the clustering method based on the possible world can be used.

4.3. The Clustering Method Based on the Possible World

In [22], the authors used an adaptive, local-structure learning method to calculate the consensus affinity matrix. In their model, the collected numerical data are used to match the probability density function (PDF) of the uncertain data. However, the authors give no algorithm for the case where the PDF is given directly. Moreover, the proposed method needs a sizable quantity of data. In this paper, Assumption 1 is proposed instead of the consistency principle.

According to Assumption 1 above, the probability of each possible world should be considered when calculating the consensus affinity matrix. Then, the objective function is shown as follows:

$$\begin{aligned} \min \sum_{j=1}^n d_{ij}^{pw} s_{ij}^{pw} + \alpha \sum_{j=1}^n s_{ij}^{pw} \\ \text{s.t. } S_i^{pw} = [s_{1i}^{pw}, s_{2i}^{pw}, \dots, s_{ni}^{pw}]^T \\ (S_i^{pw})^T \cdot \mathbf{1}^{n \times 1} = 1 \\ 0 \leq s_{ij}^{pw} \leq 1 \end{aligned} \quad (16)$$

where, d_{ij}^{pw} is a kind of distance function between O_i^{pw} and O_j^{pw} , and $S_i^{pw} = [s_{1i}^{pw}, s_{2i}^{pw}, \dots, s_{ni}^{pw}]^T$ is the normalized distance matrix for one of the possible worlds (pw).

Moreover, let the effective results of S_i

$$t = \sum_{j=1}^n \text{sgn}(s_{ij}^{pw}) \quad (17)$$

According to the conclusion of [22], t can be adjusted by α , and the optimization result is

$$s_{ij}^{pw} = \frac{1}{t} + \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^t d_{is}^{pw}}{t} - d_{ij}^{pw} \right) \quad (18)$$

where $D_i'^{pw} = [d_{1i}'^{pw}, d_{2i}'^{pw}, \dots, d_{ni}'^{pw}]^T$ is another order of D_i^{pw} , and it ranges from small to large.

According to the formulas above, the extra information about classes is required to confirm t . It is set as $t = N$ if there is no extra information about classes. That is:

$$s_{ij}^{pw} = \frac{1}{n} + \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n d_{is}'^{pw}}{n} - d_{ij}^{pw} \right) \quad (19)$$

Finally, an optimization normalized distance matrix S^* is needed for clustering the training set, which is satisfied by the following optimal model:

$$\begin{aligned} \min E \left(\|S - S^{pw}\|_F^2 \right) \\ \text{s.t. } (S_i)^T \cdot \mathbf{1}^{n \times 1} = 1 \\ 0 \leq s_{ij} \leq 1 \end{aligned} \quad (20)$$

where $S_i = [s_{1i}, s_{2i}, \dots, s_{ni}]^T$ and $S = [S_1, S_2, \dots, S_n]^T$.

According to the object function (20),

$$E(\|S - S^{pw}\|_F^2) = E\left(\sum_{i=1}^n \sum_{j=1}^n (s_{ij} - s_{ij}^{pw})^2\right) = \sum_{i=1}^n \sum_{j=1}^n E(s_{ij} - s_{ij}^{pw})^2 \quad (21)$$

On the other hand, according to (19), we have

$$s_{ij} - s_{ij}^{pw} = s_{ij} - \frac{1}{n} - \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n d_{is}^{pw}}{n} - d_{ij}^{pw} \right). \quad (22)$$

Therefore,

$$E(s_{ij} - s_{ij}^{pw})^2 = E\left(s_{ij} - \frac{1}{n} - \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n d_{is}^{pw}}{n} - d_{ij}^{pw} \right)\right)^2 \quad (23)$$

According to the properties of expectation and variance:

$$E(X^2) = E^2(X) + D(X), \quad (24)$$

$$E(aX + b) = aE(X) + b \quad (25)$$

and

$$D(aX + b) = a^2 \cdot D(X), \quad (26)$$

Equation (23) can be reduced to:

$$E(s_{ij} - s_{ij}^{pw})^2 = \left(s_{ij} - \frac{1}{n} - \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n E(d_{is}^{pw})}{n} - E(d_{ij}^{pw}) \right) \right)^2 + \frac{1}{4\alpha^2} D\left(\frac{\sum_{s=1}^n d_{is}^{pw}}{n} - d_{ij}^{pw} \right) \quad (27)$$

Obviously, (7) is equivalent to the following optimal model:

$$\min \sum_{i=1}^n \sum_{j=1}^n \left(s_{ij} - \frac{1}{n} - \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n E(d_{is}^{pw})}{n} - E(d_{ij}^{pw}) \right) \right)^2. \quad (28)$$

The optimal solution for the above optimal model can be obtained easily, which is

$$s_{ij} = \frac{1}{n} + \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n E(d_{is}^{pw})}{n} - E(d_{ij}^{pw}) \right), i = 1, 2, \dots, n. \quad (29)$$

Now, another understanding for a possible world is presented. Let us review the definition of possible world. The construction of an uncertain dataset and its PDF $f(pw)$ are known. Then, if the dimensions of the dataset are finite, which is assumed to be $\{o_{ij}\}_{j=1}^n$, the edge probability density function (EPDF) for i th dimension is:

$$f_i(O_i) = \int_{D_{pw}/D_i} f(pw) d(pw/O_i) \quad (30)$$

Moreover, if the dimensions of $O_i (i = 1, 2, \dots, n)$ are finite, which is assumed to be $\{o_{ij}\}_{j=1}^n$, the edge probability density function (EPDF) for j th dimension of $O_i (i = 1, 2, \dots, n)$ is:

$$f_{ij}(o_{ij}) = \int_{D_i/D_{ij}} f_i(O_i) d(O_i/o_{ij}) \quad (31)$$

Here, it is assumed that $distance(O_i, O_j)$ is the distance between the random variables O_i and O_j . Then, the consensus affinity matrix S can be obtained according to the following formula:

$$\begin{aligned} \min \sum_{j=1}^n d_{ij} s_{ij} + \alpha \sum_{j=1}^n s_{ij} \\ \text{s.t. } S_i = [s_{1i}, s_{2i}, \dots, s_{ni}]^T \\ (S_i)^T \cdot \mathbf{1}^{n \times 1} = 1 \\ 0 \leq s_{ij} \leq 1 \end{aligned} \quad (32)$$

where $d_{ij} = g(distance(O_i, O_j))$, and $S = [S_1, S_2, \dots, S_n]^T$.

Then, according to the analysis above, if there is no extra information about classes, the optimal solution for the object function (15) is:

$$s_{ij} = \frac{1}{n} + \frac{1}{2\alpha} \left(\frac{\sum_{s=1}^n d_{is}}{n} - d_{ij} \right). \quad (33)$$

Compared with (12), the distribution is used instead of the expectation of point distance. Therefore, (12) is appropriate for the possible world that includes fewer and simpler random variables, while (16) is appropriate for the possible world with complexity random variables in theory.

So far, when the distance-based function is confirmed, the optimization consensus affinity matrix S for the all possible worlds can be worked out.

According to the calculations above, the closer two data objects are, the larger s_{ij} is. Therefore, the value of s_{ij} may have no use when $s_{ij} < p$ (distance threshold). Then, the matrix S may need to be pruned to remove the meaningless s_{ij} . This pruning is divided into two steps: removing and normalization. In the removing step, the meaningless values are replaced by 0. In the normalization step, the meaningful value is recalculated to keep the equation:

$$\sum_{i=1}^n s_{ij} = 1, j = 1, 2, \dots, n. \quad (34)$$

The following Algorithm 1 shows the processing of pruning:

Algorithm 1 for Matrix Pruning:

Input: the matrix $S \in R^{n \times n}$ and pruning threshold p
 The processing:
 Removing step:
 For $i = 1$ to n
 For $j = 1$ to n
 If $s_{ij} < p$
 $s_{ij} = 0$
 End if
 End for
 End for
 Normalization step:
 For $i = 1$ to n
 $sum_i = \sum_{j=1}^n s_{ji}$
 For $j = 1$ to n
 $s_{ji} = \frac{s_{ji}}{sum_i}$
 End for
 End for

Moreover, in spectral analysis, if a nonnegative affinity matrix S is given, the corresponding Laplacian matrix L_s can be calculated as $L_s = D_s - \frac{S^T + S}{2}$, where D_s is a diagonal matrix and its i th diagonal element is $\sum_{j=1}^n \frac{s_{ij} + s_{ji}}{2}$. The Laplacian matrix L_s has an important property as follows [24].

Theorem 1. Let S be a nonnegative affinity matrix; then, the multiplicity k of the eigenvalue 0 of the Laplacian matrix L_s is equal to the number of connected components in the graph associated with the affinity matrix S .

It is assumed that the eigenvalues of the Laplacian matrix L_s , which is $\{\sigma_i\}_{i=1}^n$, are ordered from small to large. According to the properties of the Laplacian matrix L_s , we have the following conclusion:

$$0 = \sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_n. \quad (35)$$

If the number of clusters k is unknown, the threshold Th is set to decide k , which satisfies:

$$\sigma_k \leq Th \leq \sigma_{k+1}. \quad (36)$$

Finally, the eigenvectors of eigenvalues σ_1 to σ_k comprise the matrix $U \in R^{n \times k}$. The k -means clustering algorithm is used to cluster the row of matrix U . The clustering result is that of the training set. The Algorithm 2 for processing S is shown as follows.

Algorithm 2 for processing S :

Input: the matrix $S \in R^{n \times n}$ and clustering threshold Th
 The processing:
 $L_s = D_s - \frac{S^T + S}{2}$
 $\{\sigma_i\}_{i=1}^n$ is the set of eigenvalues of L_s
 $0 = \sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_n$.
 $\{v_i\}_{i=1}^n$ is the set of eigenvectors of L_s
 If
 $\sigma_r \leq Th \leq \sigma_{r+1}$
 $k = r$
 End if
 $U = [v_1, v_2, \dots, v_k]$

Cluster the row of matrix U according to the k-means method. These are also the clustering results for training set. Therefore, the cluster $\{C_i\}_{i=1}^k$, and the number of cluster members $\{n_i\}_{i=1}^k$ are obtained.

4.4. Updating

After clustering the training set, the data in the test set should be put into the clusters determined above. Firstly, the test set is given as follows:

The Test Set: $O = \{O_i\}_{i=n}^{n+p}$, and $O_i = [o_{1i}, o_{2i}, \dots, o_{mi}]^T$ is the data object.

The clustering updating algorithm for the test set is divided into two steps: clustering and updating. The details are shown in the following Algorithm 3:

Algorithm 3 for Clustering Updating:

Input: the center of each cluster $\{C_i\}_{i=1}^k$, and the number of cluster members $\{n_i\}_{i=1}^k$ of training set and the test set $O = \{O_i\}_{i=n}^{n+p}$.

The processing:

Clustering step:

$$\{C'_i\}_{i=1}^k = \{C_i\}_{i=1}^k.$$

For $i = n + 1$ to $n + p$

$$[d_{ij}]_{j=1}^k, d_{ij} = \text{distance}(O_i, C_j)$$

$$[d'_{ij}]_{j=1}^k, d'_{ij} = \text{distance}(O_i, C'_j)$$

$$\text{cluster}_i = \underset{j}{\operatorname{argmin}} d_{ij} \quad \text{cluster}'_i = \underset{j}{\operatorname{argmin}} d'_{ij}$$

If $\text{cluster}_i = \text{cluster}'_i$

O_i belongs to cluster_i .

Else if

$$\frac{d_{i, \text{cluster}_i}}{d_{i, \text{cluster}'_i}} \geq \frac{d'_{i, \text{cluster}'_i}}{d'_{i, \text{cluster}_i}}$$

O_i belongs to $\text{cluster}'_i$

Else

O_i belongs to cluster_i

End if

End if

End for

Centers updating step:

For $i = n + 1$ to $n + p$

If O_i belongs to cluster_j

$$C'_{\text{cluster}_i} = \frac{n_{\text{cluster}_i} C'_{\text{cluster}_i} + O_i}{n_{\text{cluster}_i} + 1}$$

$$n_{\text{cluster}_i} = n_{\text{cluster}_i} + 1$$

End if

End for

5. Simulations

In this section, comparisons with three state-of-the-art uncertain data clustering algorithms are conducted on real benchmark datasets. Moreover, an uncertain dataset that obeys the multivariate Gaussian distribution is generated, and the parameters in the PWFEM model are discussed.

In the comparisons, six common real benchmark datasets, which came from '<http://archive.ics.uci.edu/ml/>', are employed for the simulation; their details are shown in Table 1:

Table 1. Details of the adoptive datasets [25].

Dataset	Objects	Attributes	Classes
Iris	150	4	3
Wine	178	13	3
Glass	214	9	6
Ecoli	327	7	5
Waveform	5000	21	3
PhishingData [26]	1353	9	3

These datasets were originally established as collections of data with determinate values. Then, we followed the method in [27] to generate uncertainty in these datasets, and the generation method is shown as follows Algorithm 4:

Algorithm 4 The Generation Method from Numerical Data to Uncertain Data (Gaussian Type).

Input: the numerical data $a = [a_1, a_2, \dots, a_n]^T$ and the standard deviation of each attribute $[\sigma_1, \sigma_2, \dots, \sigma_n]$

Output: the corresponding uncertain data $ua = [ua_1, ua_2, \dots, ua_n]^T$

For $i = 1$ to n

$x = \text{random}, 0 < x \leq 1$

$ua_i = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a_i)^2}{2\sigma^2}}$

End for

5.1. The Clustering Accuracy

In this part, 2 widely used evaluation metric, which are accuracy (ACC) and Normalized mutual information (NMI), are adopted to compare the different clustering algorithms. In this part, the proposed clustering algorithms, PWFEM-nd and PWFEM-pd, are compared with three state-of-the-art uncertain data clustering algorithms: UK-means [26], REP [27] and PWCLU. Each clustering algorithm was run 100 times. The maximum, minimum, mean value, and variance of the ACC were calculated with respect to each algorithm. The comparisons were simulated for two cases. Case 1 is the real mean value with variance known, while case 2 is the finite measurement results, which obey the given PDF instead.

In order for the proposed model to be executed properly, the exact values of expectation and covariance need to be known. However, the datasets used in this simulation do not give those values. Therefore, the approximate values were calculated instead according to the following formula:

$$E = X \text{ and } Cov = Cov(X). \quad (37)$$

where $X = \{x_i\}_{i=1}^n$ is the dataset.

The comparisons of ACC for each algorithm in case 1 are shown in Table 2.

Table 2. Accuracy (ACC) for each algorithm in case 1.

		UK-Means	REP	PWCLU	PWFEM-nd	PWFEM-pd
Iris	Max	0.8800	0.8133	0.8133	0.8133	0.8533
	Min	0.5533	0.5533	0.5400	0.4867	0.5200
	Mean	0.7244	0.6994	0.6869	0.7181	0.7602
	Variance	0.0022	0.0021	0.0016	0.0017	0.0028
Wine	Max	0.7022	0.7022	0.5730	0.7079	0.9607
	Min	0.7022	0.7022	0.5730	0.6966	0.3202
	Mean	0.7022	0.7022	0.5730	0.6989	0.8999
	Variance	0	0	0	0	0.0173
Glass	Max	0.8333	0.7619	0.8618	0.7905	0.9286
	Min	0.2476	0.6000	0.2571	0.2286	0.3095
	Mean	0.7239	0.7078	0.7588	0.6489	0.7818
	Variance	0.0191	0.0010	0.0204	0.0173	0.0537
Ecoli	Max	0.5327	0.4953	0.5374	0.5234	0.5421
	Min	0.3458	0.2056	0.4065	0.3318	0.4299
	Mean	0.4422	0.4025	0.4905	0.4527	0.4634
	Variance	0.0012	0.0035	0.0011	0.0014	0.0009
Waveform	Max	0.5291	0.4006	0.7003	0.5199	0.7156
	Min	0.3180	0.2324	0.3945	0.4006	0.4801
	Mean	0.4350	0.3177	0.5403	0.4445	0.5706
	Variance	0.0014	0.0013	0.0025	0.0006	0.0038
PhishingData	Max	0.5639	0.4560	0.5647	0.5188	0.6061
	Min	0.4664	0.3585	0.4568	0.4508	0.4797
	Mean	0.5183	0.4218	0.5027	0.4910	0.5719
	Variance	0.0005	0.0004	0.0004	0.0002	0.0010

As shown in Table 2, in the datasets of wine and glass, the PWFEM-nd shows the best performance with maximum, minimum, and mean values. Unfortunately, it shows the worst performances with those values in the datasets of iris, Ecoli and PhishingData. As for the proposed PWFEM-pd, it shows the best performances with maximums in all datasets except wine and glass.

According to their respective algorithms, there may be plenty of reasons for the results above. Some analyses that have high probabilities are presented next.

Firstly, it is important to note that the UK-means, REP, PWCLU, and PWFEM-nd use the mean value only. Therefore, their variance values are zeros, which means the clustering results never change throughout the 100 iterations. Only PWFEM-pd uses the variance of uncertain data.

Secondly, UK-means clusters the dataset directly, while REP, PWCLU, PWFEM-nd, and PWFEM-pd cluster the dataset indirectly. Here, REP, PWCLU, PWFEM-nd, and PWFEM-pd use the model based on the possible world. Moreover, PWCLU uses the Euclidean distance ($\|\cdot\|_2$). PWFEM-nd uses the cosine similarity. PWFEM-pd uses the Kullback–Leibler divergence. Compared with the PWCLU, PWFEM-nd combines the distributions of each component in a datum. Moreover, PWFEM-pd calculates the distance in distributions directly, while PWCLU and PWFEM-nd transform the distributions into some special numbers (mean value and variance). Therefore, the clustering accuracy of PWFEM-pd may be higher than that of PWCLU in most cases. Moreover, PWFEM-pd can be regarded as having different covariances obtained randomly to that of clustering. If a covariance close to the true covariance is acquired, a high accuracy of clustering is gained.

For a clearer view of the changing of clustering accuracy with different covariances, see Figure 1.

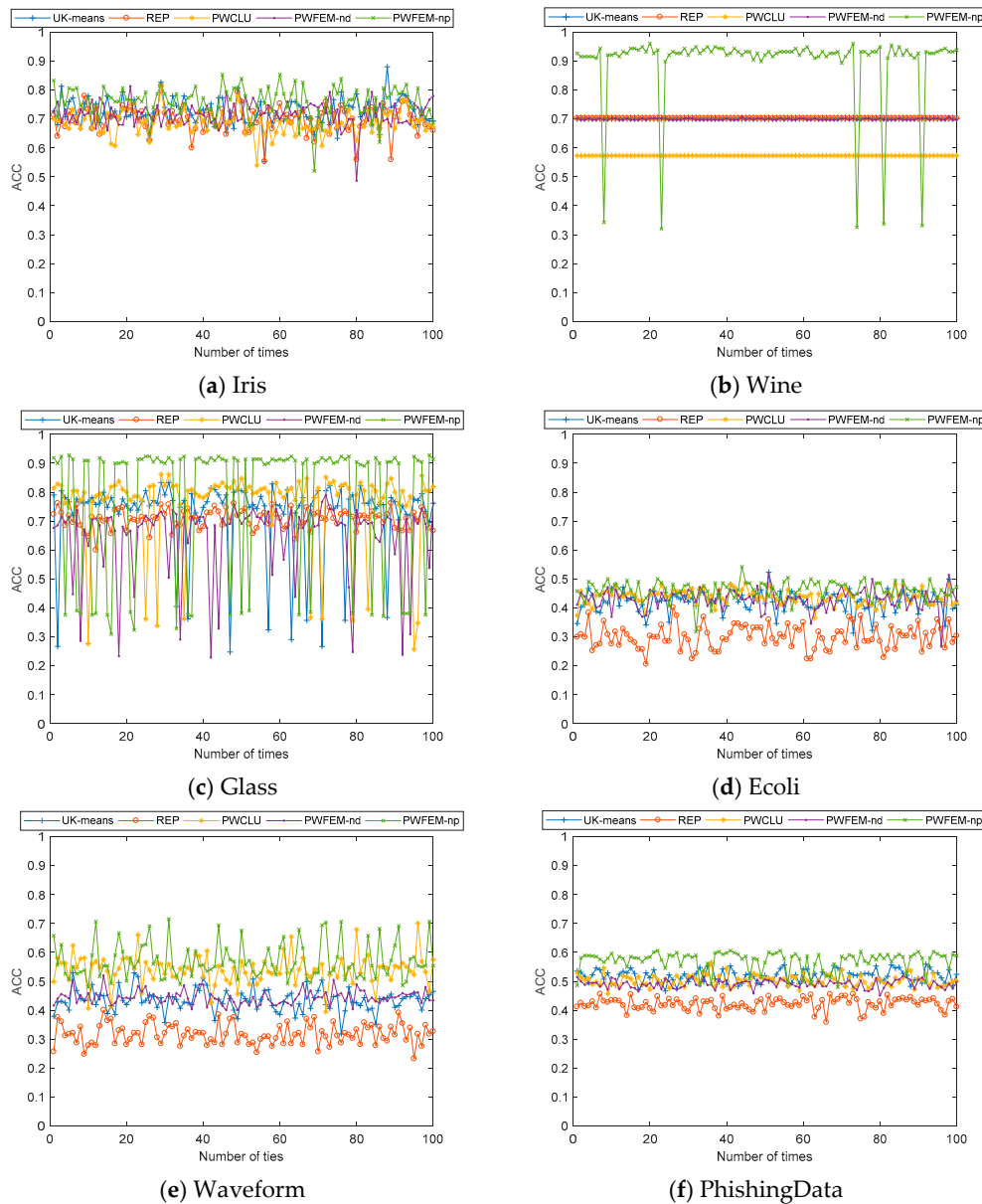


Figure 1. ACC with different clustering algorithms for 100 iterations in case 1.

As shown in Figure 1, the ACC of PWFEM-pd is sensitive to the covariance of the uncertain data. On the other hand, the impacts caused by covariances from different datasets lead to different results. Obviously, in Figure 1a,c,d,f, the ACC is highly dependent on the covariance. In Figure 1e, the ACC is divided into two parts: one is around 0.51 and the other is around 0.34, when different covariances are given. Moreover, in Figure 1b, the ACC is stable around 0.5 most times with the changing of covariance.

According to the analysis above, only for the proposed models, which are PWFEM-nd and PWFEM-pd, the ACC is sensitive to covariance. Then, the changing of mean values is added to the simulations. Therefore, the simulation results are given for case 2, which uses the generation method proposed in the beginning of this section; the results of case 2 are shown in Table 3 and Figure 2.

Table 3. NMI for each algorithm.

		UK-Means	REP	PWCLU	PWFEM-nd	PWFEM-pd
Iris	Max	0.7854	0.6809	0.6716	0.6700	0.7396
	Min	0.2694	0.3898	0.2871	0.2213	0.3162
	Mean	0.5374	0.5245	0.4834	0.5295	0.5927
	Variance	0.0050	0.0027	0.0031	0.0033	0.0054
Wine	Max	0.4946	0.4946	0.3184	0.5389	0.9551
	Min	0.4946	0.4946	0.3184	0.5136	0.3146
	Mean	0.4946	0.4946	0.3184	0.5209	0.8803
	Variance	0	0	0	0	0.0198
Glass	Max	0.7001	0.6171	0.7522	0.6288	0.8671
	Min	0.0997	0.4028	0.1643	0.0320	0.2233
	Mean	0.5511	0.5250	0.6094	0.4258	0.6911
	Variance	0.0196	0.0019	0.0223	0.0204	0.0672
Ecoli	Max	0.6544	0.6544	0.7064	0.7125	0.7309
	Min	0.3731	0.3731	0.5199	0.4679	0.3639
	Mean	0.4988	0.4988	0.6354	0.5629	0.5569
	Variance	0.0050	0.0050	0.0034	0.0054	0.0079
Waveform	Max	0.3247	0.2548	0.4645	0.3104	0.4895
	Min	0.1195	0.1286	0.1282	0.1919	0.2545
	Mean	0.2112	0.1909	0.3244	0.2392	0.3558
	Variance	0.0017	0.0008	0.0022	0.0005	0.0025
PhishingData	Max	0.2517	0.1636	0.2416	0.2200	0.3190
	Min	0.1559	0.0594	0.1452	0.1313	0.1804
	Mean	0.2088	0.1050	0.1880	0.1760	0.2803
	Variance	0.0004	0.0004	0.0005	0.0003	0.0008

As shown in Table 3, when combining the maximum value and minimum value, the clustering results of all clustering methods change. This means all the clustering methods are sensitive to the mean value. Moreover, the sensitivity to each clustering method varies. Obviously, the fluctuation ranges of all clustering methods in iris and glass are the most drastic. On the other hand, the clustering accuracy of the PWFEM-pd algorithm is always higher than that of the PWFEM-nd, but its stability is lower than that of the PWFEM-nd. Besides, compared with Table 2, the NMI are lower than the ACC for the same dataset, which means that in the clustering results of the model, the accuracy of each class is inconsistent, with some categories having high precision and some having low precision.

For a clearer view of the changing of clustering accuracy with different covariances and mean values, see Figure 2.

As shown in Figure 2, the PWFEM-pd has a similar fluctuation as that shown in Figure 1. Unfortunately, this clustering method is sensitive to both mean value and covariance. Therefore, it is hard to distinguish the main reason. Next, the remaining four clustering methods are discussed.

Firstly, similar to the conclusion in Table 2, Figure 2c,d in all clustering methods show a drastic fluctuation. For UK-means and CK-means, they show a drastic fluctuation in Figure 2a and are stable in Figure 2b,e,f. For PWCLU, it is stable in Figure 2a,b,e,f. For PWFEM-nd, it is stable in Figure 2a,b,f, while it is stable at two ranges in Figure 2e.

According to the analysis above, the situations of the proposed methods are clearer. However, the variation tendency with the mean value and covariance are not clear. Therefore, a specific dataset was generated to investigate the above issues.

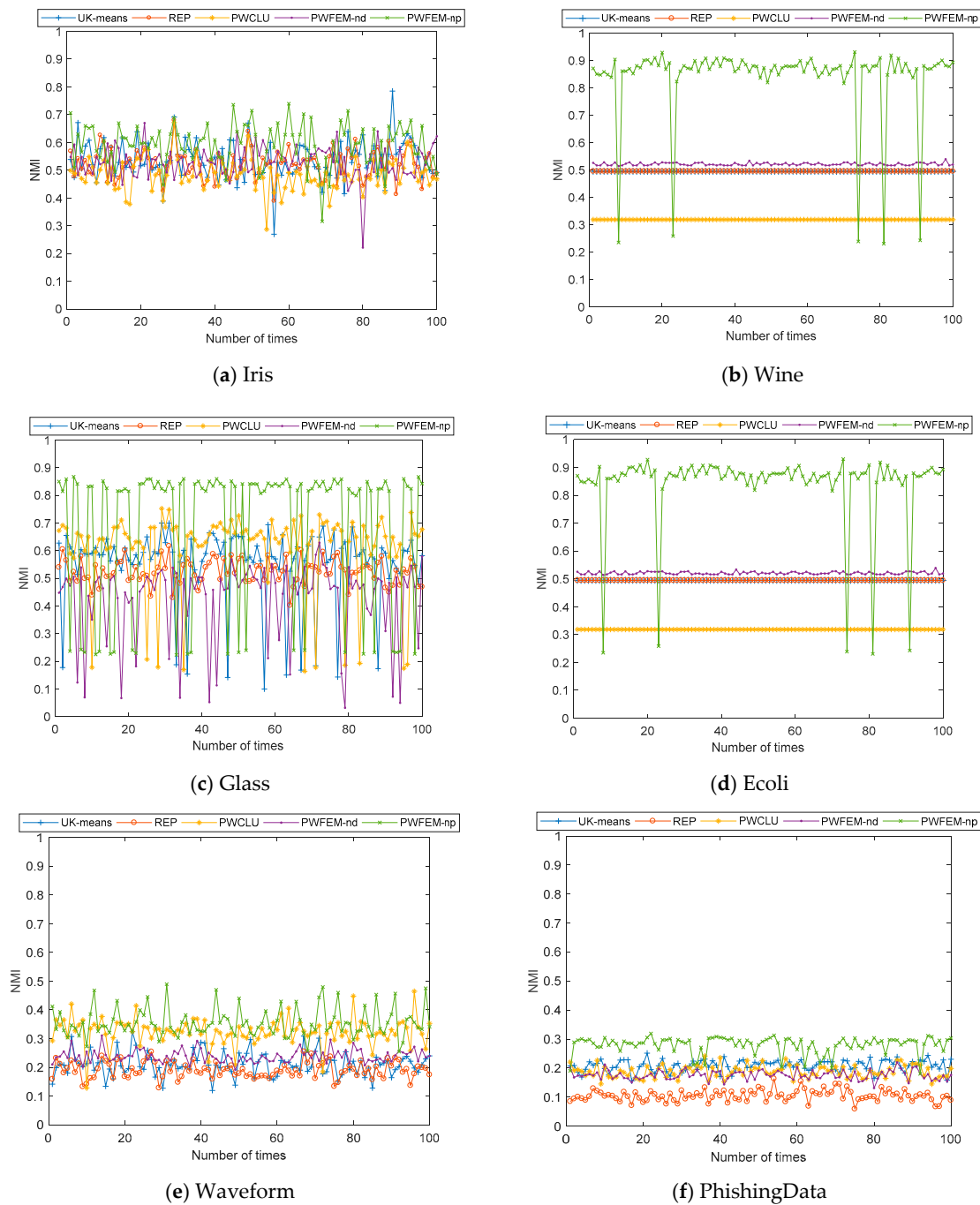


Figure 2. NMI with different clustering algorithms for 100 iterations in case 2.

5.2. The Simulation with a Specific Dataset

In this part, a specific dataset is generated to analyze the impacts of mean value and covariance. The generated dataset consisted of two dimensions, and the number of data points was set at 1000. It was divided into three clusters, whose centers were $[0, 0]$, $[100, 0]$, and $[0, 100]$. The distance between the datum and its center was randomly distributed in $[0, r]$. The variance for each dimension was $\sigma_i (i = 1, 2)$. Moreover, it was set as $\sigma_1 = \sigma_2 = \sigma$. The correlation coefficient of these two dimensions was ρ . Therefore, the covariance of this dataset was:

$$\begin{bmatrix} \sigma & \rho\sigma \\ \rho\sigma & \sigma \end{bmatrix}.$$

Next, the parameters r , σ , and ρ are discussed.

In this simulation, $\sigma = 2$, $\rho = 0$, and r was from 1 to 100. As shown in Figure 3, the ACCs of all methods were 1 before about 50, and then reduced with increasing r . This simulation results are in accordance with common sense.

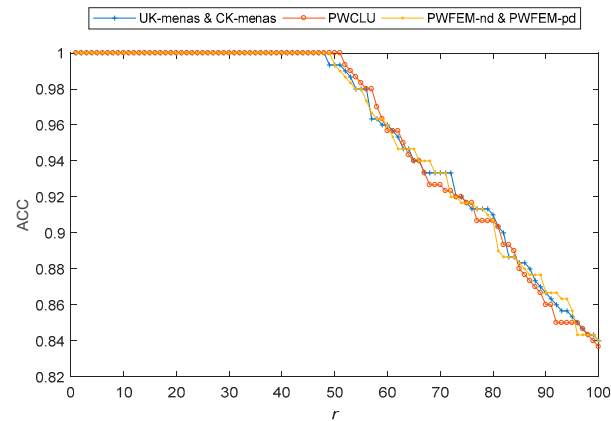


Figure 3. Change of ACCs with r values from 1 to 100.

On the other hand, if $\rho = 0$ and r is fixed, σ can vary the distance between the data evenly. Therefore, it cannot affect the clustering results, and the simulation proves it. Because the ACC curves of all methods are lines parallel to the X-axis, the figure was omitted.

Finally, the simulation for ρ is discussed with $\sigma = 2$, $r = 20, 40, 60$, and 80 , and $-1 < \rho < 1$. The simulation results are shown in Figure 4.

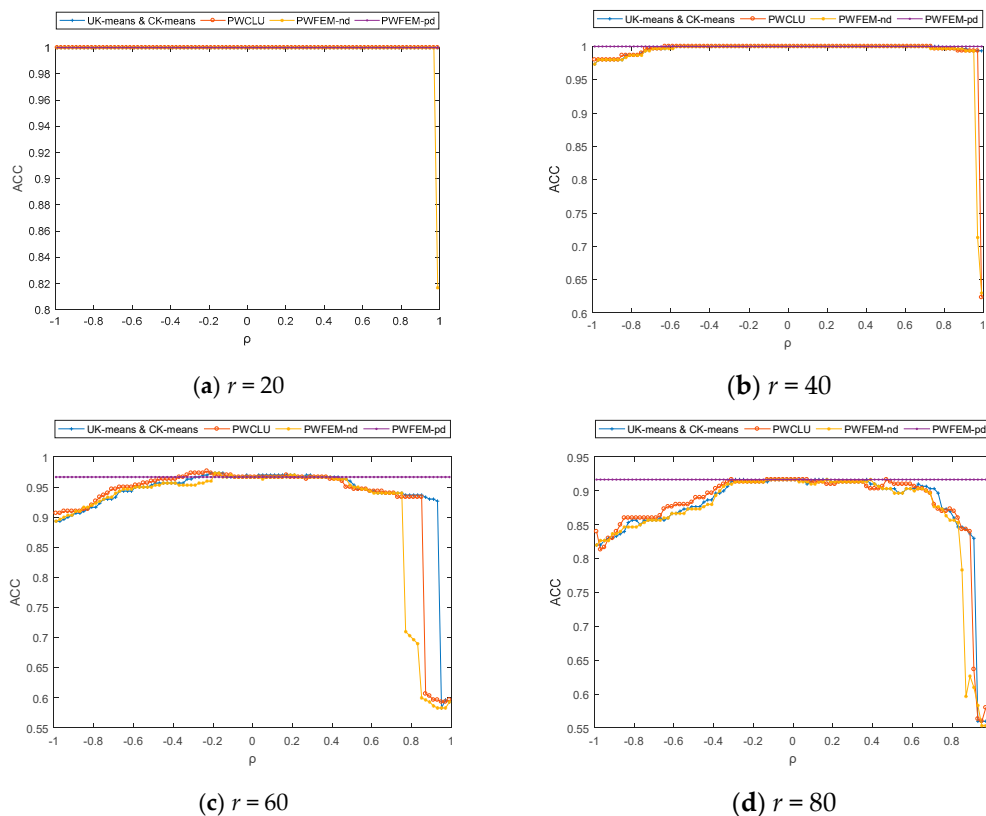


Figure 4. Change of ACCs with ρ from -1 to 1 .

As shown in Figure 4, when $r < 50$, the ACCs are stable for all methods with $-1 < \rho < 1$. This is because the cluster structure is prominent in this condition, whereas the effect of ρ on the clustering result is weak. Moreover, when $r > 50$, the ACCs of UK-means, CK-means, PWCLU, and PWFEM-nd show significant changes between in $(-1, -0.7)$ and in $(0.7, 1)$. In these two intervals, ρ makes the data points even messier. Therefore, the ACCs of clustering results decrease if the data points are not processed. On the other hand, the ACCs become stable when $-0.7 < \rho < 0.7$. Obviously, the effect of ρ on the clustering results is weak.

6. Conclusions

In this paper, a possible world-based fusion estimation model for uncertain data is proposed. It includes two methods, which are the PWFEM-nd and PWFEM-pd. The PWFEM-nd is based on a data perspective, which uses a bottom-up method to cluster the data. The PWFEM-pd uses clustering according to the uncertain data directly. Both these methods depend more on the probability density distribution of uncertain data. We performed some simulations and confirmed that the proposed methods showed better performance in terms of probabilistic accuracy. The accuracy is highly dependent on the accuracy of covariance.

The discussion in the last section is incomplete. Obviously, it gets more complex when dimension increases. Only some simple conclusions are given in the simulation. In addition, the exact covariance is not usually obtained in actual scenarios. In any case, the proposed methods provide a new way to treat uncertain data clustering. The issues mentioned above are also to be addressed in future works.

Author Contributions: C.L.: Manuscript writing and data analysis. Z.Z.: Algorithm research and design, manuscript revising. W.W.: Algorithm for discussion. H.-C.C.: Algorithm for discussion. X.L.: The data collection and manuscript revising. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant number 2018YFC0831304) and the National Natural Science Foundation of China (grant number 61772064).

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in "<http://archive.ics.uci.edu/ml/index.php>", reference numbers are [25,26].

Conflicts of Interest: The authors declare no conflict of interest.

References

- Delias, P.; Doumpos, M.; Grigoroudis, E.; Manolitzas, P.; Matsatsinis, N. Supporting healthcare management decisions via robust clustering of event logs. *Knowl. Based Syst.* **2015**, *84*, 203–213. [CrossRef]
- Gaglio, S.; Re, G.L.; Morana, M. Human Activity Recognition Process Using 3-D Posture Data. *IEEE Trans. Hum. Mach. Syst.* **2017**, *45*, 586–597. [CrossRef]
- Waller, L.A.; Turnbull, B.W.; Clark, L.C.; Nasca, P. Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics* **1992**, *3*, 281–300. [CrossRef]
- Matthews, G.; Warm, J.S.; Shaw, T.H.; Finomore, V.S. Predicting battlefield vigilance: A multivariate approach to assessment of attentional resources. *Ergonomics* **2014**, *57*, 856–875. [CrossRef] [PubMed]
- Sun, W.; Yuan, D.; Ström, E.G.; Brännström, F. Cluster-Based Radio Resource Management for D2D-Supported Safety-Critical V2X Communications. *IEEE Trans. Wirel. Commun.* **2015**, *15*, 1. [CrossRef]
- Zagouras, A.; Pedro, H.T.C.; Coimbra, C.F.M. Clustering the solar resource for grid management in island mode. *Sol. Energy* **2014**, *110*, 507–518. [CrossRef]
- Li, M.; Xu, D.; Zhang, D.; Zou, J. The seeding algorithms for spherical k-means clustering. *J. Glob. Optim.* **2019**, *76*, 695–708. [CrossRef]
- Lu, H.; Zhang, R.; Li, S.; Li, X. Spectral Segmentation via Midlevel Cues Integrating Geodesic and Intensity. *IEEE Trans. Cybern.* **2013**, *43*, 2170–2178. [CrossRef]

9. Sokoloski, S. Implementing a Bayes Filter in a Neural Circuit: The Case of Unknown Stimulus Dynamics. *Neural Comput.* **2017**, *29*, 2450–2490. [[CrossRef](#)]
10. Sinopoli, B.; Schenato, L.; Franceschetti, M.; Poolla, K.; Jordan, M.I.; Sastry, S.S. Kalman filtering with intermittent observations. *IEEE Trans. Autom. Control* **2004**, *49*, 1453–1464. [[CrossRef](#)]
11. Zhang, W.; Zhang, Z.; Zeadally, S.; Chao, H.C.; Leung, V. CMASM: A Multiple-algorithm Service Model for Energy-delay Optimization in Edge Artificial Intelligence. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4216–4224. [[CrossRef](#)]
12. Zhang, W.; Zhang, Z.; Wang, L.; Chao, H.C.; Zhou, Z. Extreme learning machines with expectation kernels. *Pattern Recognit.* **2019**, *96*, 1–13. [[CrossRef](#)]
13. Chau, M.; Cheng, R.; Kao, B.; Ng, J. Uncertain data mining: An example in clustering location data. In Proceedings of the Advances in Knowledge Discovery and Data Mining 10th Pacific-Asia Conference, Singapore, 9–12 April 2006.
14. Kriegel, H.P.; Pfeifle, M. Hierarchical density-based clustering of uncertain data. In Proceedings of the Data Mining Fifth IEEE International Conference, Houston, TX, USA, 27–30 November 2005.
15. Volk, P.B.; Rosenthal, F.; Hahmann, M.; Habich, D.; Lehner, W. Clustering Uncertain Data with Possible Worlds. In Proceedings of the Proceedings of the 25th International Conference on Data Engineering, Shanghai, China, 29 March–2 April 2009.
16. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
17. Garcia, E.; Hausotte, T.; Amthor, A. Bayes filter for dynamic coordinate measurements Accuracy improvement, data fusion and measurement uncertainty evaluation. *Meas. J. Int. Meas. Confed.* **2013**, *46*, 3737–3744. [[CrossRef](#)]
18. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
19. Costa, P.J. Adaptive model architecture and extended Kalman-Bucy filters. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 525–533. [[CrossRef](#)]
20. Julier, S.J.; Uhlmann, J.K.; Durrant-Whyte, H.F. A New Approach for Filtering Nonlinear Systems. In Proceedings of the American Control Conference, Seattle, DC, USA, 21–23 June 1995.
21. Haykin, S. *Kalman Filtering and Neural Networks*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
22. Liu, H.; Zhang, X.; Zhang, X. Possible World-based consistency learning model for clustering and classifying uncertain data. *Neural Netw.* **2018**, *102*, 48–66. [[CrossRef](#)]
23. Sinkkonen, J.; Kaski, S. Clustering Based on Conditional Distributions in an Auxiliary Space. *Neural Comput.* **2014**, *14*, 217–239. [[CrossRef](#)]
24. Luxburg, U.V. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
25. Dua, D.; Graff, C. *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml/>]; University of California, School of Information and Computer Science: Irvine, CA, USA, 2019.
26. Abdelhamid, N.; Ayesh, A.; Thabtah, F. Phishing detection based Associative Classification data mining. *Expert Syst. Appl.* **2014**, *41*, 5948–5959. [[CrossRef](#)]
27. Gullo, F.; Tagarelli, A. Uncertain centroid based partitional clustering of uncertain data. *Proc. VLDB Endow.* **2012**, *5*, 610–621. [[CrossRef](#)]