

A REVIEW ON DIMENSIONALITY REDUCTION USING COPULA APPROACH IN DATA MINING

Sumaiya Maryam* and Sriram Yadav

M.Tech. Scholar, MITS, Bhopal.

A.P., CSE, MITS Bhopal.

Article Received on 18/08/2019

Article Revised on 08/09/2019

Article Accepted on 29/09/2019

***Corresponding Author**

Sumaiya Maryam

M.Tech. Scholar, MITS,
Bhopal.

ABSTRACT

Copula approach is a Sampling-based dimensionality reduction technique. Removing linearly redundant combined dimensions, giving a convenient way to generate correlated multivariate random variables.

Managing the integrity of the original information, deducting the dimension of data space without losing valuable information. The modern trends in collecting very large and diverse datasets have created a great challenge in data analysis. The recent trends in collecting very large and diverse datasets have created a great challenge in data analysis. One of the attributes of these gigantic datasets is that they often have significant amounts of redundancies. The use of very large multi-dimensional data will result in more noise, redundant data, and the possibility of unconnected data entities. To efficiently manage data represented in a high-dimensional space and to address the impact of redundant dimensions on the final results, a new technique has been proposed for the dimensionality reduction using Copulas and the LU-decomposition (Forward Substitution) method. The proposed method is compared favorably with existing approaches on real-world datasets: Diabetes, Waveform, two versions of Human Activity Recognition based on Smartphone, and Thyroid Datasets taken from machine learning repository in terms of dimensionality reduction and efficiency of the method, which are performed on statistical and classification measures.

KEYWORDS: Data mining, Integrity, Sampling, Copulas, Dimensionality reduction, Redundancy, LU-Decomposition.

INTRODUCTION

Dimensionality reduction technique is based on probabilistic and sampling models; therefore, one needs to recall some fundamental concepts. These include the notions of a Probability Density Function (PDF), Cumulative Distribution Function (CDF), a random variable used to generate samples from a probability distribution, a Copula to model dependencies of data without imposing constraints to specific types of marginal probability density functions, and dependence and rank correlations of multivariate random variables to measure dependencies of the dimensions.

The following table gives basic notations used throughout this paper

Table 1: Basic notations.

Primitive	Definition
X	$n \times m$ data matrix (random variable).
X^i	i^{th} row of the matrix X .
X_j	j^{th} column of the matrix X .
$F_j(.)$	CDF of the j^{th} column.
$f_j(.)$	PDF of the j^{th} column.
C	Gaussian Copula of the matrix X .
c	Density associated with C .
C_{ij}	Empirical Copula of the matrix X .
Σ	Correlation matrix of C .
X^t	Transposed matrix of X .
v_{ij}	Value of the i^{th} row and j^{th} column.

Let f be the Probability Density Function (PDF) of a random variable X . The probability distribution of X consists in calculating the probability $P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m)$, for all $(X_1, \dots, X_m)$ belongs to R_m . It is completely specified by the CDF F which is defined in (Rubinstein & Kroese, 2011) as follows:

$$F(x_1, x_2, \dots, x_m) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m) \text{ -----(1)}$$

1.1 Random Variable Generation

The problem of generating a sample from a one-dimensional cumulative distribution function CDF by calculating the inverse transform sampling. To illustrate the problem, let X be a continuous random variable with a CDF $F(x) = P[X \leq x]$, and U be a continuous uniform distribution over the interval $[0, 1]$. The transform $X = F^{-1}(U)$ denotes the inverse transform sampling function of a given continuous uniform variable $U = F(X)$ in $[0, 1]$, where $F^{-1}(u) = \{ \min x, F(x) \geq u \}$ (Rubinstein & Kroese, 2011). So the simple steps used for generating a sample $X \sim F$ are given as follows (Rubinstein & Kroese, 2011):

1. Generate $U \sim U[0, 1]$;
2. Return $X = F^{-1}(U)$.

The usual problem is how to combine one-dimensional distribution functions to form multivariate distributions and how to estimate and simulate their density $f(x_1, x_2, \dots, x_m)$ to obtain the required number of random samples of $X_i, i=1, \dots, m$, especially in high-dimensional spaces. This problem will be explained in the following section.

1.2 Modeling with Copulas

The first usage of Copulas is to provide a convenient way to generate correlated multivariate random variable distributions and to present a solution for the difficulties of transformation of the density estimation problem.

To illustrate the problem of invertible transformations of m -dimensional continuous random variables $X_1, \dots, X_m$ according to their *CDF*, into m independently uniformly-distributed variables $U_1 = F_1(X_1), U_2 = F_2(X_2), \dots, U_m = F_m(X_m)$, let $f(x_1, x_2, \dots, x_m)$ be the probability density function of $X_1, \dots, X_m$, and let $c(u_1, u_2, \dots, u_m)$ be the joint probability density function of $U_1, U_2, \dots, U_m$. In general, the estimation of the probability density function $f(x_1, x_2, \dots, x_m)$ can provide a nonparametric form (unknown families of distributions). In this case, the probability density function $c(u_1, u_2, \dots, u_m)$ of $U_1, U_2, \dots, U_m$ has been estimated instead of that $X_1, \dots, X_m$ to simplify the density estimation problem, and then simulated it to achieve the random samples $X_1, \dots, X_m$ using the inverse transformations $X_i = F^{-1}(U_i)$.

Sklar's Theorem showed that there exists a unique m -dimensional Copula C in $[0, 1]^m$ with standard uniform marginal distributions $U_1, \dots, U_m$. (Nelsen, 2007) states that every distribution function F with margins $F_1, \dots, F_m$ can be written as:

$$F(X_1, \dots, X_m) = C(F_1(X_1), \dots, F_m(X_m)), \forall (X_1, \dots, X_m) \in \mathbb{R}^m \quad (2)$$

To evaluate the suitability of a selected Copula with estimated parameter and to avoid the introduction of any assumptions on the distribution $F_i(X_i)$, one can utilize an empirical CDF of a marginal $F_i(X_i)$, to transform m samples of X into m samples of U . An empirical Copula is useful for examining the dependence structure of multivariate random vectors. Formally, the empirical Copula is given by the following equation:

$$C = 1/m \left(\sum_{k=1}^n I(v_k \mathbf{j} \leq v \mathbf{i} \mathbf{j}) \right) \quad (3)$$

Where the function $I_{(arg)}$ is the indicator function, which equals 1 if arg is true and 0 otherwise. Here, m is used to keep the empirical CDF less than 1, where m is the number of observations. In the following, we will focus on the Copula those results from a standard multivariate Gaussian Copula.

1.3 Gaussian Copula

The difference between the Gaussian Copula and the joint normal CDF is that the Gaussian Copula allows having different marginal CDF types from the joint distribution (Nelsen, 2007). However, in probability theory and statistics, the multivariate normal distribution is a generalization of the one-dimensional normal distribution. The Gaussian Copula is defined as follows:

$$C(\Phi(x_1), \dots, \Phi(x_m)) = 1/\sqrt{|\Sigma|} \exp(-1/2 X^t(\Sigma^{-1} - I)X) \quad (4)$$

Where $\Phi(x_i)$ is the CDF standard Gaussian distribution of $f_{i(x_i)}$, i.e., $X_i \sim N(0, 1)$, and Σ is the correlation matrix. The resulting Copula $C(u_1, \dots, u_m)$ is called Gaussian Copula. The density associated with $C(u_1, \dots, u_m)$ is obtained with the following equation:

$$C(u_1, \dots, u_m) = 1/\sqrt{|\Sigma|} \exp[-1/2 \xi^t (\Sigma^{-1} - I) \xi] \quad (5)$$

Where $u_i = \Phi(x_i)$, and $\xi = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m))^T$.

1.4 Dependence and Rank Correlation

Since the Copula of a multivariate distribution describes its dependence structure, it might be appropriate to use measures of dependence which are Copula-based. The Pearson correlation measures the relationship $\Sigma = \text{cov}(X_i, X_j) / (\sigma_{X_i} \sigma_{X_j})$ where $\text{cov}(X_i, X_j)$ is the covariance of X_i and X_j while $\sigma_{X_i}, \sigma_{X_j}$ are the standard deviations of X_i and X_j . Kendall rank correlation (also known as Kendall's coefficient of concordance) is a non-parametric test that measures the strength of dependence between two random samples X_{ip}, X_{jp} of n observations. The notion of concordance can be defined by the following equation:

$$\tau = P[(X^i - X^j)(X^i_t - X^j_t) > 0] - P[(X^i - X^j)(X^i_t - X^j_t) < 0] \quad (6)$$

For the Gaussian Copula, Kendall's τ can be calculated as follows:

$$\tau = 2/\pi \arcsin \Sigma_{X_i X_j} \quad (7)$$

2. LITERATURE REVIEW

2.1 Linear dimensionality reduction

Principal Component Analysis (PCA) is a well established method for dimensionality reduction. It derives new variables (in decreasing order of importance) that are linked by linear combinations of the original variables and are uncorrelated. Several models and techniques for data reduction based on PCA have been proposed (Sasikala & Balamurugan, 2013). (Zhai et al., 2014) proposed a maximum likelihood approach to the multi-size PCA problem. The covariance based approach was extended to estimate errors within the resulting PCA decomposition. Instead of making all the vectors of fixed size and then computing a covariance matrix, they directly estimate the covariance matrix from the multi-sized data using nonlinear optimization. (Kerdprasop et al., 2014) studied the recognition accuracy and the execution times of two different statistical dimensionality reduction methods applied to the biometric image data, which are: PCA and Linear Discriminant Analysis (LDA). The learning algorithm that has been used to train and recognize the images is a support vector machine with linear and polynomial kernel functions. The main drawback of reducing dimensionality with PCA is that it can only be used if the original variables are correlated, and homogeneous, if each component is guaranteed to be independent and if the dataset is normally distributed. If the original variables are not normalized, PCA is not effective.

The Sparse Principal Component Analysis (SPCA) (Zou et al., 2006) is an improvement of the classical method of PCA to overcome the problem of correlated variables using the LASSO technique. LASSO is a promising variable selection technique, producing accurate and sparse models. SPCA is based on the fact that PCA can be written as a regression problem where the response is predicted by a linear combination of the predictors. Therefore, a large number of coefficients of principal components become zero, leading to a modified PCA with sparse loading. Many studies on data reduction based on SPCA have been presented. (Shen & Huang, 2008) proposed an iterative algorithm named sparse PCA via regularized SVD (sPCA-rSVD) that uses the close connection between PCA and singular value decomposition (SVD) of the data matrix and extracts the PCs through solving a low rank matrix approximation problem. (Bai et al., 2015) proposed a method based on sparse principal component analysis for finding an effective sparse feature principal component (PC) of multiple physiological signals. This method identifies an active index set corresponding to the non-zero entries of the PC, and uses the power iteration method to find the best direction.

Singular Value Decomposition (SVD) is a powerful technique for dimensionality reduction. It is a particular case of the matrix factorization approach and it is therefore also related to PCA. The key issue of an SVD decomposition is to find a lower dimensional feature space by using the matrix product $U S V$, where U and V are two orthogonal matrices and S is a diagonal matrix with $m \times m$, $m \times n$, and $n \times n$ dimensions, respectively. SVD retains only $r \times n$ positive singular values of low effect to reduce the data, and thus S becomes a diagonal matrix with only r non-zero positive entries, which reduces the dimensions of these three matrices to $m \times r$, $r \times r$, and $r \times n$, respectively. Many studies on data reduction have been presented which are built upon SVD, such as the ones used in (Zhang et al., 2010) and (Watcharapinchai et al., 2009). (Lin et al., 2014) developed a dimensionality reduction approach by applying the sparsified singular value decomposition (SSVD). Their paper demonstrates how SSVD can be used to identify and remove nonessential features in order to facilitate the feature selection phase, to analyze the application limitations and the computational complexity. However, the application of SSVD on large datasets showed a loss of accuracy and makes it difficult to compute the eigenvalue decomposition of a matrix product $A^T A$, where A is the matrix of the original data.

2.2 Nonlinear dimensionality reduction

A vast literature devoted to nonlinear techniques has been proposed to resolve the problem of dimensionality reduction, such as manifold learning methods, e.g., Locally Linear Embedding (LLE), Isometric mapping (Isomap), Kernel PCA (KPCA), Laplacian Eigenmaps (LE), and a review of these methods is summarized in (Gisbrecht & Hammer, 2015; Wan et al., 2016). KPCA (Kuang et al., 2015) is a nonlinear generalization of PCA in a high-dimensional kernel space constructed using kernel functions. By comparing with PCA, KPCA computes the principal eigenvectors using the kernel matrix, rather than the covariance matrix. A kernel matrix is done by computing the inner product of the data points. LLE (Hettiarachchi & Peters, 2015) is a nonlinear dimensionality reduction technique based on simple geometric intuitions. This algebraic approach computes the low-dimensional neighborhood preserving embeddings. The neighborhood is preserved in the embedding based on a minimizing cost function in input space and output space, respectively. Isomap (Zhang et al., 2016) explores an underlying manifold structure of a dataset based on the computation of geodesic manifold distances between all pairs of data points. The geodesic distance is determined as the length of the shortest path along the surface of the manifold between two data points. It first constructs a neighborhood graph between all data points based on the connection of each point to all its

neighbors in the input space. Then, it estimates geodesic distances of all pairs of points by calculating the shortest path distances in the neighborhood graph. Finally, multidimensional scaling (MDS) is applied to the arising geodesic distance matrix to find a set of low-dimensional points that greatly match such distances.

2.3 Sampling dimensionality reduction

Other widely used techniques are based on sampling. They are used for selecting a representative subset of relevant data from a large dataset. In many cases, sampling is very useful because processing the entire dataset is computationally too expensive. In general, the critical issue of these strategies is the selection of a limited but representative sample from the entire dataset. Various random, deterministic, density biased sampling, pseudo-random number generator and sampling from non-uniform distribution strategies exist in the literature (Rubinstein & Kroese, 2011). However, very little work has been done on the Pseudo-random number generator and sampling from non-uniform distribution strategies, especially in the multi-dimensional case with heterogeneous data. Naïve sampling methods are not suitable for noisy data which are part of real-world applications, since the performance of the algorithms may vary unpredictably and significantly. The random sampling approach effectively ignores all the information present in the samples which are not part of the reduced subset (Whelan et al., 2010). An advanced data reduction algorithm should be developed in multi-dimensional real-world datasets, taking into account the heterogeneous aspect of the data. Both approaches (Colomé et al., 2014)(Fakoor & Huber, 2012) are based on sampling and a probabilistic representation from uniform distribution strategies. The authors of (Fakoor & Huber, 2012) proposed a method to reduce the complexity of solving Partially Observable Markov Decision Processes (POMDP) in continuous state spaces. The paper uses sampling techniques to reduce the complexity of the POMDPs by reducing the number of state variables on the basis of samples drawn from these distributions by means of a Monte Carlo approach and conditional distributions. The authors in (Colomé et al., 2014) applied dimensionality reduction to a recent movement representation used in robotics, called Probabilistic Movement Primitives (ProMP), and they addressed the problem of fitting a low-dimensional, probabilistic representation to a set of demonstrations of a task. The authors fitted the trajectory distributions and estimated the parameters with a model-based stochastic using the maximum likelihood method. This method assumes that the data follow a multivariate normal distribution which is different from the typical assumptions about the relationship between the empirical data. The best we can do is to examine the sensitivity of

results for different assumptions about the data distribution and estimate the optimal space dimension of the data.

2.4 Similarity measure dimensionality reduction

There are other widely used methods for data reduction based on similarity measures (Wencheng, 2010) (Pirolla et al., 2012) (Zhang et al., 2010). According to (Dash et al., 2015), the presence of redundant or noisy features degrades the classification performance, requires huge memory, and consumes more computational time. (Dash et al., 2015) proposes a three-stage dimensionality reduction technique for microarray data classification using a comparative study of four different classifiers, multiple linear regression (MLR), artificial neural network (ANN), k -nearest neighbor (k -NN), and naïve Bayesian classifier to observe the improvement in performance. In their experiments, the authors reduce the dimension without compromising the performance of such models. (Deegalla et al., 2012) proposed a dimensionality reduction method that employs s classification approaches based on the k -nearest neighbor rule. The effectiveness of the reduced set is measured in terms of the classification accuracy. This method attempts to derive a minimal consistent set, i.e., a minimal set which correctly classifies all the original samples (Whelan et al., 2010). (Venugopalan et al., 2014) discussed the ongoing work in the field of pattern analysis for bio-medical signals (cardio-synchronous waveform) using a Radio Frequency Impedance Interrogation (RFII) device for the purpose of user identification. They discussed the feasibility of reducing the dimensions of these signals by projecting them into various sub-spaces while still preserving inter-user discriminating information, and they compared the classification performance using traditional dimensionality reduction methods such as PCA, independent component analysis (ICA), random projections, or k -SVD-based dictionary learning. In the majority of cases, the authors see that the space obtained based on classification carries merit due to the dual advantages of reduced dimension and high classification.

Developing effective clustering methods for high-dimensional datasets is a challenging task (Whelan et al., 2010). (Boutsidis et al., 2015) studied the topic of dimensionality reduction for k -means clustering that encompasses the union of two approaches: 1) A feature selection-based algorithm selects a small subset of the input features and then the k -means is applied on the selected features. 2) A feature extraction-based algorithm constructs a small set of new artificial features and then the k -means is applied on the constructed features. The first feature extraction method is based on random projections and the second is based on fast

approximate SVD factorization. (Sun et al., 2014) developed a tensor factorization based on a clustering algorithm (k-mean), referred to as Dimensionality Reduction Assisted Tensor Clustering (DRATC). In this algorithm, the tensor decomposition is used as a way to learn low-dimensional representation of the given tensors and, simultaneously, clustering is conducted by coupling the approximation and learning constraints, leading to the PCA Tensor Clustering and Non-negative Tensor Clustering models.

Problems identification and Objectives

The most serious problem is the presence of missing values in datasets. Missing values can result in loss of efficiency of the dimensionality reduction approach, lead to complications in handling and analyzing the data, or distort the relationship between the data distribution. Also, it is interesting to investigate the possibility of using Meta heuristics or hybrid approaches to determine a solution of the proposed optimization problem in the Big Data setting. Objectives of the research work are to overcome several problems identified and strong efforts to improve the performance of the dimensionality reduction approach in very large datasets.

Proposed Methodology

The approach presented in this paper for dimensionality reduction in very large datasets is based on the theory of Copulas and the LU-decomposition method (Forward Substitution). The main goal of the method is to reduce the dimensional spaces of data without losing important/interesting information. On the other hand, the goal is to estimate the multivariate joint probability distribution without imposing constraints on specific types of marginal distributions of dimensions. Figure 1 shows an overview of the proposed reduction method which operates in two main steps.

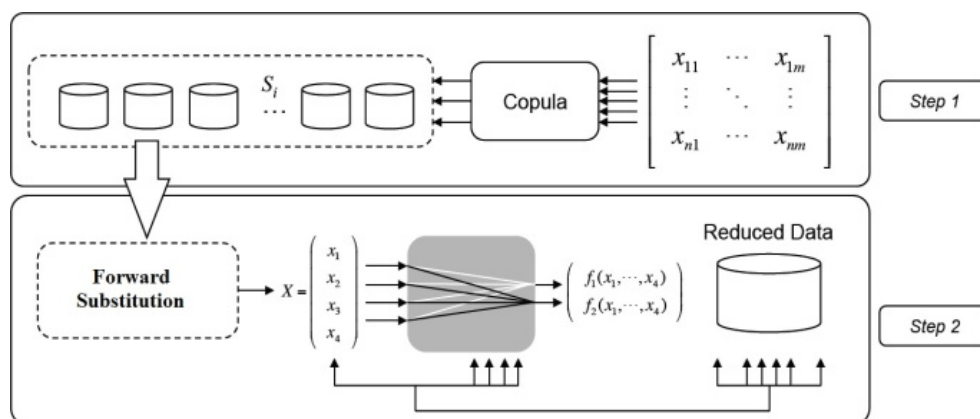


Figure 1: Overview of the proposed reduction method.

In the first step, large raw datasets are decomposed into smaller subsets when calculating the data dependencies using a Copula by taking into account heterogeneous data and removing the data which are strongly dependent. In the second step, we want to reduce the space dimensions by eliminating dimensions that are linear combinations of others. Then we will find the coefficients of the linear combination of dimensions by applying the LU-decomposition method (Forward Substitution) to each subset to obtain an independent set of variables in order to improve the efficiency of data mining algorithms. The two different steps of the proposed method are as follows (See also Figure 1):

Step 1: Construction of dependent sample subset S_i , ($i=1,...,kt$)

In order to decompose the real-world dataset into smaller dependent sample subsets, vectors are considered which are linearly dependent in the original data.

Empirical Copula will be calculated first to better observe the dependencies between variables. According to the marginal distributions from the observed and approved empirical Copula, we can determine the theoretical.

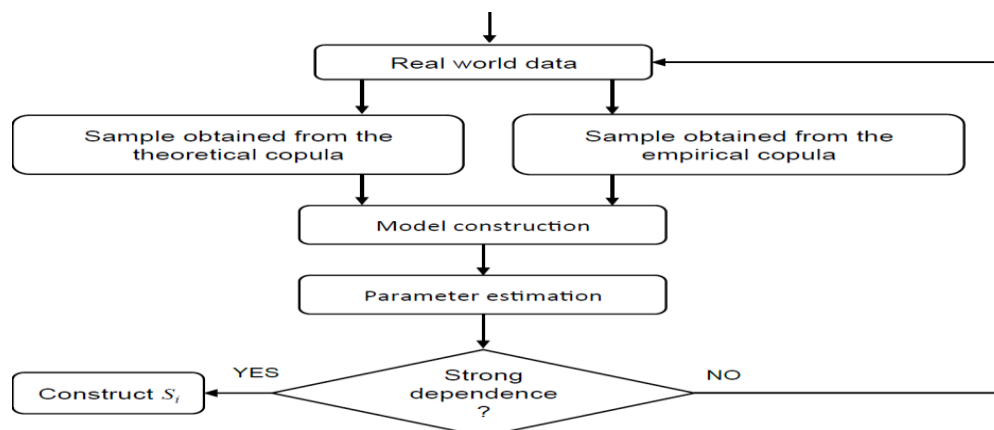


Figure 2: Construction of the subsets S_i , ($i=1,...,kt$).

Copula, that links univariate marginal distributions to their joint multivariate distribution function, and then we will regroup dimensions having the strong correlation relationship in each sample subset $S_{i,(i=1,...,kt)}$ by estimating the parameters of the Copula. In this paper, we have presented the Gaussian Copula that corresponds to our experimental results. An illustration of the Copula method is given in Figure 2.

The dependence between two continuous random variables X_1 and X_2 is defined as follows: If the correlation parameter ρ is greater than 0.5, then X_1 and X_2 are positively correlated,

meaning that the values of X_1 increase as the values of X_2 increase (i.e., the more each attribute implies the other). Hence, a higher value may indicate that X_1 and X_2 are positively dependent, and probably have a highly redundant attribute, then these two samples will be made as in the same subset $S_{i,(i=1,\dots,k)t}$. When the parameter of the Copula ρ of the two continuous random variables X_1 and X_2 is greater than 0.7, then X_1 and X_2 have a strong dependence. If the resulting value is equal or less than 0, then X_1 and X_2 are independent and there is no correlation between them.

The output of the sample subset $S_{i,(i=1,\dots,k)t}$ represents a matrix that retains only dependent samples of the original matrix in order to detect, and remove a maximum of the redundant dimensions, which are linear combinations of others, in the second step.

Step 2: LU-decomposition method

The key idea behind the use of the Forward Substitution method is to solve the linear system equations as given by the samples $S_{i,(i=1,\dots,k)t}$ with an upper-triangular coefficient matrix in order to find the coefficients of linear sample combinations and to provide a low linear space ($X_i; i=1, \dots, k$) of the original matrix as shown in Figure 3.

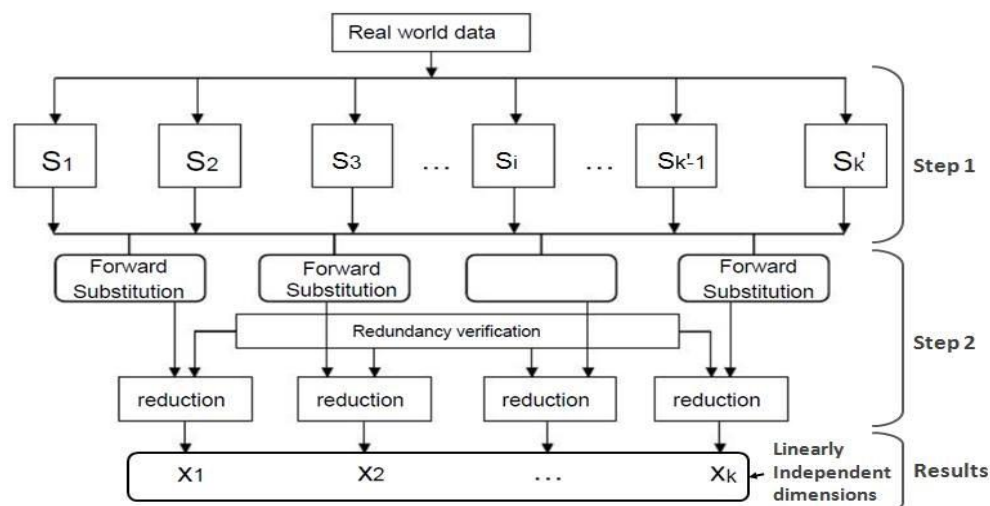


Figure 3: Schema of dimensionality reduction.

The LU decomposition method is an efficient procedure for solving a system of linear equations $\alpha^T \hat{X} = C$, and it can help accelerate the computation. When C is a column vector in the dependent sample subsets $S_{i,(i=1,\dots,k)t}$, and α_j is an output vector representing the relationship between dimensions or the coefficients of the linear combination of dimensions,

$\mathbf{S}_i; (i=1, \dots, k'-1)$ induces a lower triangular matrix without column C . We conclude that each matrix $\hat{S}_{i, (i=1, \dots, k'-1)}$ induces a lower triangular matrix of the following form:

$$(\mathbf{S}) \{ \alpha_{11} = c_1$$

$$\{ \alpha_{12} + \alpha_{22} = c_2 \quad (8)$$

$$\{ \alpha_{1n} + \alpha_{2n} + \dots + \alpha_{nn} = c_n$$

From the above equations, we see that $\alpha_1 = c_1/x_{11}$. Thus, we compute α_1 from the first equation and substitute it into the second to compute $\alpha_2, \dots$, etc. Repeating this process, we reach equation i , $2 \leq i \leq n$, using the following formula:

$$\alpha_i = 1/x_i [c_i - \sum_{j=1}^{i-1} \alpha_{ij} x_{ij}], \quad i=2, \dots, n \quad (9)$$

Algorithm 1: Dimensionality linear combination reduction method

Input: Vector C and a lower triangular matrix $\mathbf{S}$;

Output: Vector α .

Begin

$$\alpha_1 = c_1/x_{11}$$

for $i := 2$ to n do

$$\alpha_i = c_i$$

for $j := 1$ to $i-1$ do

$$\alpha_i = \alpha_i - x_{ij}\alpha_j$$

end

$$\alpha_i = \alpha_i / x_{ii}$$

end

The algorithm 1 used for this resolution makes $(n \times (n-1))/2$ additions and subtractions, $(n \times (n-1))/2$ multiplications and n divisions to calculate the solution, a global number of operations in the order of n^2 .

CONCLUSIONS

In this paper, we have proposed a new method for dimensionality reduction in the data pre-processing phase of mining high-dimensional data. This approach is based on the theory of Copulas (sampling techniques) to estimate the multivariate joint probability distribution without constraints of specific types of marginal distributions of random variables that represent the dimensions of our datasets. A Copula based model provides a complete and

scale- free description of dependency that is thereafter used to detect the redundant values. A more extensive evaluation is made by eliminating dimensions that are linear combinations of others after having decomposed the data, and using the LU- decomposition method. We have reformulated the problem of data reduction as a constrained optimization problem. We have compared the proposed approach with well-known data mining methods using five real-world datasets taken from the machine learning repository in terms of the dimensionality reduction and the efficiency of the methods. The efficiency of the proposed method was improved by using the both statistical and classification methods. The different results obtained show the effectiveness of our approach which outperforms significantly the performance of dimensionality reduction comparing to other methods, i.e., it provided a smaller bias with more better standard deviations, a highest precision, and a lowest recall with all classifiers for all databases.

REFERENCES

1. Agatonovic-Kustrin, S., & Beresford, R. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, 2000; 22: 717–727.
2. Bai, D., Liming, W., Chan, W., Wu, Q., Huang, D., & Fu, S. Sparse principal component analysis for feature selection of multiple physiological signals from flight task. In *Control, Automation and Systems (ICCAS)*, 15th International Conference on, 2015; 627–631.
3. Boutsidis, C., Zouzias, A., Mahoney, M. W., & Drineas, P. Randomized dimensionality reduction for-means clustering. *Information Theory, IEEE Transactions on*, 2015; 61: 1045–1062.
4. Colomé, A., Neumann, G., Peters, J., & Torras, C. Dimensionality reduction for probabilistic movement primitives. In *Humanoid Robots (Humanoids)*, 2014 14th IEEE-RAS International Conference on, 2014; 794– 800.
5. Dash, R., Misra, B., Dehuri, S., & Cho, S.-B. Efficient microarray data classification with three-stage dimensionality reduction. In *Intelligent Computing, Communication and Devices*, 2015; 805–812.
6. Deegalla, S., Boström, H., & Walgama, K. Choice of dimensionality reduction methods for feature and classifier fusion with nearest neighbor classifiers. In *Information Fusion (FUSION)*, 2012 15th International Conference on, 2012; 875–881.

7. Derrac, J., Chiclana, F., Garcia, S., & Herrera, F. Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets. *Information Sciences*, 2016; 329: 144–163.
8. Fakoor, R., & Huber, M. A sampling-based approach to reduce the complexity of continuous state space POMDPs by decomposition into coupled perceptual and decision processes. In *Machine Learning and Applications (ICMLA)*, 11th International Conference on, 2012; 687–692.
9. Gisbrecht, A., & Hammer, B. Data visualization by nonlinear dimensionality reduction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2015; 5: 51–73.
10. Han, J., & Kamber, M. (San Francisco). *Data Mining: Concepts and Techniques*. 13: 978-1-55860-901-3 (2nd ed.). Diane Cerra, 2006.
11. Hettiarachchi, R., & Peters, J. Multi-manifold LLE learning in pattern recognition. *Pattern Recognition*, 2015; 48: 2947–2960.
12. Houari, R., Bounceur, A., & Kechadi, M.-T. A new method for dimensionality reduction of multi-dimensional data using copulas. In *Programming and Systems (ISPS)*, 11th International Symposium on, 2013; 40–46.
13. Houari, R., Bounceur, A., & Kechadi, T. A New Approach for Dimensionality Reduction of Large Multi-Dimensional Data Based on Sampling Methods for Data Mining, 2013.
14. Houari, R., Bounceur, A., Kechadi, T., Tari, A., & Euler.
15. R. A new method for estimation of missing data based on sampling methods for data mining. In *Advances in Computational Science, Engineering and Information Technology*, 2013; 89–100.
16. Kerdprasop, N., Chanklan, R., Hirunyawanakul, A., & Kerdprasop, K. An empirical study of dimensionality reduction methods for biometric recognition. In *Security Technology (SecTech)*, 2014 7th International Conference on, 2014; 26–29.
17. Kuang, F., Zhang, S., Jin, Z., & Xu, W. A novel svm by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection. *Soft Computing*, 2015; 1–13.
18. Lichman, M. Uci machine learning repository, university of california, irvine, school of information and computer sciences, <http://archive.ics.uci.edu/ml/>, 2013.
19. Lin, P., Zhang, J., & An, R. Data dimensionality reduction approach to improve feature selection performance using sparsified svd. In *Neural Networks (IJCNN)*, 2014 International Joint Conference on, 2014; 1393–1400.

20. Nelsen, R. B. An introduction to copulas. (2nd ed.). Springer Science & Business Media, 2007.
21. Pirolla, F. R., Felipe, J., Santos, M. T., & Ribeiro, M. X. Dimensionality reduction to improve content- based image retrieval: A clustering approach. In Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on (pp. 752–753). IEEE. Rubinstein, R. Y., & Kroese, D. P. (2011). Simulation and the Monte Carlo method volume 707. John Wiley & Sons. Saoudi, M., Bounceur, A., Euler, R., & Kechadi, T. (2016). Data mining techniques applied to wireless sensor networks for early forest fire detection. In Proceedings of the International Conference on Internet of things and Cloud Computing, 2012; 71.
22. Sasikala, S., & Balamurugan, S. A. A. Data classification using pca based, 2013.