A Comparative Study of Different Dimensionality Reduction Methods with Naïve Bayes Classifier for Mapping Customer Requirements to Product Configurations

Yao Jiao[1] and Yu Yang[2*]

Department of Industrial Engineering, College of Mechanical Engineering,
Chongqing University, Chongqing, China, 400044
[1]jiaoyao@cqu.edu.cn, [2]yuyang@cqu.edu.cn

Abstract

Mapping customer requirements to product configurations are difficult due to the uncertainty and ambiguity of customers' expression. The Naïve Bayes Classifier (NBC) is suitable to quantify the expression of customers, and to map their requirements to configurations with good performance. However, the prerequisite of manually independent of product attributes for NBC require preprocess. Dimensionality reduction methods are effective for simplifying the data complexity while separating the correlations between data Against the background, this paper conducts a comparative study of 7 dimensionality reduction methods as preprocess procedure for integrating with NBC to map customer requirements to product configurations. Two realistic design cases are illustrated for the comparison, and the outcomes are measured by the accuracy and F-measure. The results of this study imply several findings that the loss of information has great impact on all methods, and linear methods are more sensitive to the loss of information, and several nonlinear methods are more capable in handling the loss of information than other methods, and local linear methods are suggested compared with global nonlinear methods.

Keywords: Customer Requirement; Product Configuration; Naïve Bayes Classifier; Dimensionality Reduction

1. Introduction

Many researches have been conducted to elicit and understand customer needs depend on customers' explicit specifications on product [1]. It requires customers to have domain knowledge about product to map their requirements to design parameters. However, customers may find the process stressful and unpleasant due to the lack of expertise [2]. As a probabilistic classifier, Naïve Bayes Classifier (NBC) is suitable to quantify the natural uncertainty and ambiguity of customer requirements [3]. NBC is based on the strong assumption that the attributes are independent with each other [4].

Meanwhile, on one hand, customers are becoming more and more knowledgeable and fastidious on product [5]. It demands product to satisfy customers with providing divergent and various attributes. On other hand, more areas are considered in product design process, e.g., engineering, business and art [6]. It expands the extensions of attributes for measuring product. These two reasons result in complexity and high dimensionality of product attributes data. High dimensional data contains intertwined relationships and dependencies between attributes that may cause NBC inefficient and fallible, eventually results in the curse of dimensionality [7].

Dimensionality reduction is an extremely useful way to avoid the curse of dimensionality by extracting independent and principle variables while retaining the

---

* Corresponding author

original variation of original data [8]. The post low-dimensional representations after reducing dimensionality are typically independent that suitable for NBC input. Though many available reduction methods have been proposed and applied, several questions are still urgent to be discussed, for instance, (*a*) which method should be combined with NBC for mapping customer requirements to product configurations; (*b*) Are nonlinear methods performs better in real design works?

In this study, we conduct a comparative study to analyze the differentiation of 7 popular dimensionality methods integrating with NBC for mapping customer requirements to product configurations based on 2 realistic product design cases. The differences are compared on the basis of 2 measures. This study would like to make contributes to the research of practical application and theoretical development of dimensionality reduction methods in product design issues.

The rest of the paper is organized as follows. The next section introduces the NBC and dimensionality reduction methods. In the section of methodology, the cases and data are given out, along with the analysis steps and the performance measures. The results section shows the results and discussions of analysis, and this paper summarizes the contributions in the conclusion section.

2. Related Works

2.1 Naïve Bayes Classifier

NBC is a simple and efficient probabilistic method based on Bayes rules, which is suitable to handle uncertain information. Also, the fast convergence speed makes the small training cases practicable in the design issues, e.g., mapping customer requirements to configuration plans [1].

Let the set $\{R_1, R_2, …, R_p\}$ denotes customer requirements on a product attributes, where R_k denotes a requirement. To map customer requirements to the optimal configuration plan, $P\left(\overline{U} \mid R_1, R_2, …, R_p\right) = \max\left\{\prod P\left(\overline{X} \mid R_1, R_2, …, R_p\right)\right\}$ is needed, where $\overline{U}$ denotes a configuration plan, $\overline{X}$ denotes a product attributes. Assumed all product attributes are conditionally independent, then based on Bayes rules, $P\left(\overline{X} \mid R_1, R_2, …, R_p\right)$ can be calculated as:

$$P\left(\overline{X} \mid R_1, R_2, …, R_p\right) = \frac{P\left(\overline{X}\right) P\left(R_1, R_2, …, R_p \mid \overline{X}\right)}{P\left(X_1, X_2, …, X_p\right)} \tag{1}$$

$$\propto P\left(\overline{X} \mid R_1, R_2, …, R_p\right) = P\left(\overline{X}\right) \cdot \prod_{k=1}^{p} P\left(R_k \mid \overline{X}\right)$$

This study estimates the probabilities of $P\left(\overline{X}\right)$ and $P\left(R_k \mid \overline{X}\right)$ by applying maximum likelihood estimation (MLE) in the learning stage for NBC, as:

$$P\left(\overline{X} = x_i\right) = \frac{|x_i|}{\sum_i |x_i|}, \text{ and } P\left(R_k \mid \overline{X} = x_i\right) = \frac{|R_k \cap x_i|}{|x_i|} \tag{2}$$

where $|x_i|$ denotes the number of product configuration x_i, and $|R_k \cap x_i|$ is the number of cases of the choosing of x_i. The learning stage can be applied by using existing customer requirements and selected configuration plans.

The NBC is trained to select the configuration plan with highest probability, as:

$$P\left(\overline{U} \mid R_1,\ R_2,\ \ldots,\ R_p\right) = \max\left\{\prod P\left(\overline{X} \mid R_1,\ R_2,\ \ldots,\ R_p\right)\right\}$$

$$= \arg\ \max_{\overline{X}} \prod P\left(\overline{X} \mid R_1,\ R_2,\ \ldots,\ R_p\right) \tag{3}$$

$$= \arg\ \max_{\overline{X}} \prod \left(P\left(\overline{X}\right) \cdot \prod_{k=1}^{p} P\left(R_k \mid \overline{X}\right)\right)$$

This study takes 7 prevalent dimensionality reduction methods to conduct the comparative research. These methods can be divided into linear methods, such as: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and nonlinear methods, such as: Kernel Principle Component Analysis (KPCA), Multi-Dimensional Scaling (MDS), Isomap, Local Linear Embedding (LLE), and Laplacian Eigenmaps (LE). The basic introduction and implementation steps are shown in the following sections.

2.2. Principle Component Analysis (PCA)

PCA is a traditional and the most popular linear method to reduce data dimensionality, while retaining most of the variation in the data set [9]. PCA performs competitively when compared with other methods on real world tasks [10]. PCA is summarized in the following steps.

Step 1: Calculate the covariance matrix $\mathbf{C}$ of the data set $\mathbf{X} = (X_1,\ X_2,\ \ldots,\ X_n)^{\mathrm{T}}$:

$$\mathbf{C} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)^{\mathrm{T}} \tag{4}$$

Step 2: Solve the eigenvector problem:

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \tag{5}$$

Step 3: Find the linear mapping matrix $\mathbf{M}$, of which columns are formed by eigenvectors that corresponding to the largest eigenvalues.

Step 4: Let $\mathbf{Y}$ denote the low dimensional data set, and y_i point to the high dimensional data x_i , as:

$$\mathbf{Y} = \mathbf{X}\mathbf{M} \tag{6}$$

Thus the out-of-sample extensions of PCA is straightforward as:

$$y^{new} = x^{new}\mathbf{M} \tag{7}$$

2.3. Independent Component Analysis (ICA)

ICA is supposed to find underlying factors or components from multivariate statistical data, and looks for components that are both statistically independent, and non-Gaussian [11]. Previous studies consider ICA as an effective linear feature extraction method for improving the classification performance [12]. ICA is summarized in the following steps.

Step 1: Let m elements from the random vector $\mathbf{X} = (X_1,\ X_2,\ \ldots,\ X_m)^{\mathrm{T}}$ are mixtures of m independent elements of random vector $\mathbf{S} = (S_1,\ S_2,\ \ldots,\ S_m)^{\mathrm{T}}$, and let $\mathbf{A}$ represents an mixing matrix, as:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{8}$$

Step 2: Suppose an un-mixing matrix $\mathbf{W}$ (i.e. the inverse of $\mathbf{A}$), which will give low dimensional vector $\mathbf{Y}$, the best possible approximation of $\mathbf{S}$, as:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \cong \mathbf{S} \tag{9}$$

Step 3: According to the Lagrange conditions, the optima of $E\left\{G(\mathbf{W}^{\mathrm{T}}\mathbf{X})\right\}$ under the constraint $E\left\{G(\mathbf{W}^{\mathrm{T}}\mathbf{X})^2\right\} = \|\mathbf{W}\|^2 = 1$ are obtained at points where the gradient of the Lagrangian is zero:

$$F = E\left\{\mathbf{X}g(\mathbf{W}^\mathrm{T}\mathbf{X})\right\} + \beta\mathbf{W} = 0 \tag{10}$$

where g is the derivative of the function G.

Step 4: Evaluating the gradient:

$$\frac{\partial F}{\partial \mathbf{W}} = E\left\{\mathbf{X}\mathbf{X}^\mathrm{T}g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} + \beta\mathbf{I} \tag{11}$$

where $\mathbf{I} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\mathrm{T}$, $\tilde{\mathbf{A}}$ is the transformation of $\mathbf{A}$ by whitening and is orthogonal.

Step 5: A reasonable approximation is:

$$E\left\{\mathbf{X}\mathbf{X}^\mathrm{T}g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} \approx E\left\{\mathbf{X}\mathbf{X}^\mathrm{T}\right\}E\left\{g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} = E\left\{g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\}\mathbf{I} \tag{12}$$

Thus $\mathbf{W}$ is obtained as:

$$\mathbf{W} = \mathbf{W} - \left[E\left\{\mathbf{X}g\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} + \beta\mathbf{W}\right]\Big/\left[E\left\{g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} + \beta\right] \tag{13}$$

Simplify Eq. (13) by multiplying both sides of the previous equation by $E\left\{g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} + \beta$. This gives, after straightforward algebraic simplification:

$$\mathbf{W} = E\left\{\mathbf{X}g\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\} - E\left\{g'\left(\mathbf{W}^\mathrm{T}\mathbf{X}\right)\right\}\mathbf{W} \tag{14}$$

And the out-of-sample extensions of ICA is straightforward as:

$$y^{new} = x^{new}\mathbf{W} \tag{15}$$

2.4. Kernel principle Component Analysis (KPCA)

KPCA is a nonlinear approach to generate PCA in the high dimensional data using kernel function [10]. KPCA does not involve any nonlinear optimization problem, making it simple as standard PCA [13]. Recently, KCPA has been successfully applied in many areas [14]. KPCA is summarized in the following steps.

Step 1: Suppose a kernel matrix $\mathbf{G}$ produced by high dimensional data points x_i, and the entries are defined as:

$$g_{ij} = \kappa(x_i, x_j) \tag{16}$$

where κ is the kernel function and is based on the Gauss kernel function as:

$$\kappa(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{17}$$

where σ^2 is the variance and a parameter of the Gauss kernel function.

Step 2: Modify the entries to center $\mathbf{G}$ by:

$$g_{ij} = g_{ij} - \frac{1}{N}\sum_l g_{il} - \frac{1}{N}\sum_l g_{jl} + \frac{1}{N^2}\sum_{lm} g_{lm} \tag{18}$$

Step 3: Solve the eigenvector problem:

$$\mathbf{G}\mathbf{v} = \lambda\mathbf{v} \tag{19}$$

where the columns of $\mathbf{v}$ is formed by the principal eigenvectors v_k corresponding to the largest eigenvalues λ_k.

Step 4: The low dimensional data representations y_i of the high dimensional data points x_i is given as:

$$y_{ik} = \frac{1}{\sqrt{\lambda_k}}v_{ik} \tag{20}$$

where v_{ik} is the ith element of the kth principal eigenvector v_k. Then the out-of-sample extensions of KPCA is straightforward as:

$$y_k^{new} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{N} v_{ik} \kappa^C\left(x_i, x^{new}\right) \tag{21}$$

where κ^C is the centered kernel function given by:

$$\kappa^C\left(x_i, x^{new}\right) = \kappa\left(x_i, x^{new}\right) - \frac{1}{N} \sum_l \kappa\left(x_i, x_l\right)$$
$$- \frac{1}{N} \sum_m \kappa\left(x^{new}, x_m\right) + \frac{1}{N^2} \sum_{lm} \kappa\left(x_l, x_m\right) \tag{22}$$

2.5. Multidimensional Scaling (MDS)

MDS is a nonlinear dimensionality reduction method with retaining similarity and distance between data points [10]. MDS has two types, Metric MDS and Non-metric MDS. This study takes the Metric MDS as the selected type of MDS. MDS is summarized in the following steps.

Step 1: Let $d\left(X_i, X_j\right)$ is the Euclidean distance of data points X_i and X_j, which are from data set $\mathbf{X} = (X_1, X_2, \ldots, X_N)^{\mathrm{T}}$ in D dimensional space, then:

$$d\left(X_i, X_j\right) = \left\| X_i - X_j \right\|_2 = \sqrt{\sum_{k=1}^{D} \left(x_{ki} - x_{kj}\right)^2} \tag{23}$$

Step 2: Set up squared proximity matrix:

$$\mathbf{D} = \left(d^2\left(X_i, X_j\right)\right)_{N \times N} = \left[\left\| X_i \right\|^2 - 2X_i^T X_j + \left\| X_j \right\|^2 \right] = \mathbf{B1'} - 2\mathbf{X^T X} + \mathbf{1B^T} \tag{24}$$

where $\mathbf{B} = \left(\left\| X_1 \right\|^2, \left\| X_2 \right\|^2, \ldots, \left\| X_N \right\|^2 \right)^{\mathrm{T}}$.

Step 3: Let $\mathbf{J} = \mathbf{I} - \frac{1}{N}\mathbf{11'}$, the $\mathbf{J1} = 0$, $\mathbf{J^T} = \mathbf{J}$, and $\mathbf{JX} = \left[X_i - \mu \right]$, $\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$, apply double centering:

$$\mathbf{JDJ} = \mathbf{JB1'J} - 2\mathbf{JX^T XJ} + \mathbf{J1BTJ} = -2\mathbf{X^T X} \tag{25}$$

Step 4: Suppose $\mathbf{H} = \mathbf{X^T X} = -\mathbf{JDJ}/2$, and apply Eigen decomposition as $\mathbf{H} = \mathbf{V\Lambda V^T}$. Let $V_1, V_2, \ldots, V_N \in R^N$ are the eigenvectors corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$, thus the low dimensional data set representation $\mathbf{Y}$ of high dimensional data $\mathbf{X}$ is given by:

$$\mathbf{Y} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_d})[V_1, V_2, \ldots, V_d] \tag{26}$$

2.6. Isomap

Isomap is a nonlinear dimensionality reduction method, and can be viewed as a kernel matrix [10]. It is based on replacing the Euclidean distance by an approximation of the geodesic distance on the manifold. Isomap has been applied in many areas [15]. Isomap is summarized in the following steps.

Step 1: Suppose every data point x_i is connected with its k nearest neighbors, x_i comes from the high dimensional dataset $\mathbf{X}$ in a neighborhood graph P.

Step 2: Compute the shortest path $T(x_i, x_j)$ of x_i and x_j by applying Dijkstra's or Floyd's shortest-path algorithm, which forms a good estimate of geodesic distance.

Step 3: Set up the pairwise geodesic distance matrix $\mathbf{T}$ by computing all the geodesic distances of data points in $\mathbf{X}$.

Step 4: Let $\mathbf{G} = \mathbf{T}^2$, and g_{ij} as the entries. Double center $\mathbf{G}$ by modifying g_{ij} as:

$$g_{ij} = -\frac{1}{2}\left(g_{ij} - \frac{1}{N}\sum_l g_{il} - \frac{1}{N}\sum_l g_{jl} + \frac{1}{N^2}\sum_{lm} g_{lm}\right) \tag{27}$$

Step 5: Solve the eigenvector problem as in Eq. (19). Then select m principal eigenvectors v_k $(1 \le k \le m)$ corresponding to the m largest eigenvalues λ_k $(1 \le k \le m)$, v_k forms the columns of matrix $\mathbf{V}$.

Step 6: The low dimensional data representations y_i of the high dimensional data points x_i is given as:

$$y_{ik} = \sqrt{\lambda_k}\, v_{ik} \tag{28}$$

where v_{ik} is the ith element of the kth principal eigenvector v_k. Then the out-of-sample extensions of Isomap is straightforward as:

$$y_k^{new} = \frac{1}{\sqrt{\lambda_k}}\sum_{i=1}^{N} v_{ik}\kappa^C\left(x_i, x^{new}\right) \tag{29}$$

where κ^C is the centered kernel function given by:

$$\kappa^C\left(x_i, x^{new}\right) = -\frac{1}{2}\left(\mathbf{T}^2\left(x_i, x^{new}\right) - \frac{1}{N}\sum_l \mathbf{T}^2\left(x_i, x_l\right)\right.$$
$$\left. -\frac{1}{N}\sum_m \mathbf{T}^2\left(x^{new}, x_m\right) + \frac{1}{N^2}\sum_{lm}\mathbf{T}^2\left(x_l, x_m\right)\right) \tag{30}$$

2.7. Local linear Embedding (LLE)

LLE develops a graph representation of data points like Isomap as a nonlinear dimensionality reduction method. In contrast to Isomap, LLE attempts to preserve local properties of data, thus it is less sensitive to short-circuiting than Isomap [10]. Many researches have emphasized the efficiency of LEE in reducing data dimensionality [16]. LLE is summarized in the following steps.

Step 1: Find the k nearest neighbors for data point x_i to construct a local neighborhood.

Step 2: Set up a sparse matrix $\mathbf{W}$ of local predictive weights w_{ij}, minimizing $\varepsilon(\mathbf{W})$ to best reconstruct data point from its nearest neighbors as:

$$\varepsilon(\mathbf{W}) = \sum_{i=1}^{N}\left\| x_i - \sum_{j=1}^{N} w_{ij}x_j \right\|^2 \tag{31}$$

subject to $w_{ij} = 0$.

Step 3: Suppose matrix $\mathbf{G} = (\mathbf{I} - \mathbf{W})^{\mathrm{T}}(\mathbf{I} - \mathbf{W})$, where I is the identity matrix. Then solve the eigenvector problem as in Eq. (19). Select m principal eigenvectors v_k $(1 \le k \le m)$ corresponding to the m smallest nonzero eigenvalues λ_k $(1 \le k \le m)$, v_k forms the columns of matrix $\mathbf{V}$.

Step 4: The low dimensional data representations y_i of the high dimensional data points x_i is given as:

$$y_{ik} = v_{ik} \tag{32}$$

where v_{ik} is the ith element of the kth principal eigenvector v_k. Then the out-of-sample extensions of LLE is straightforward as:

$$y_k^{new} = \frac{1}{\sqrt{\lambda_k}}\sum_{i=1}^{N} v_{ik}w\left(x_i, x^{new}\right) \tag{33}$$

2.8. Laplacian Eigenmaps (LE)

LE is similar to LLE as a nonlinear local dimensionality reduction method that attempts to attain low dimensional data by retaining local properties of the manifold. LE uses weighted manner to computer the distances between a data point and its nearest neighbors in low dimensional representation [10]. Many researches have focused on the improvement and application of LE recently [17]. LE is summarized in the following steps.

Step 1: Construct a neighborhood graph P in which every data point x_i is connected to its k nearest neighbors.

Step 2: For all points x_i and x_j in graph P that are connected by an edge, the weight of the edge is computed using the Gauss kernel function as Eq. (17), leading to a sparse adjacency matrix $\mathbf{W}$.

Step 3: Set up the degree diagonal matrix $\mathbf{R}$ of $\mathbf{W}$, the entries of $\mathbf{R}$ are row sum of $\mathbf{W}$. Then, set up the graph Laplacian $\mathbf{L}$, as $\mathbf{L} = \mathbf{R} - \mathbf{W}$. Solve the eigenvector problem as:

$$\mathbf{L}\mathbf{v} = \lambda \mathbf{R}\mathbf{v} \tag{34}$$

Then, select m principal eigenvectors v_k $(1 \le k \le m)$ corresponding to the m smallest nonzero eigenvalues λ_k $(1 \le k \le m)$, v_k forms the columns of matrix $\mathbf{V}$.

Step 5: The low dimensional data representations y_i of the high dimensional data points x_i is given as Eq. (32). Then the out-of-sample extensions of LE is straightforward as:

$$y_k^{new} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{N} v_{ik} \overline{\kappa}\left(x_i, x^{new}\right) \tag{35}$$

where $\overline{\kappa}$ is a normalized kernel function as given by:

$$\overline{\kappa}\left(x_i, x^{new}\right) = \frac{1}{N} \frac{\kappa(x_i, x^{new})}{\sqrt{\frac{1}{N}\sum_l \kappa(x_i, x_l) \times \frac{1}{N}\sum_m \kappa(x^{new}, x_m)}} \tag{36}$$

where κ is the Gauss kernel function as given by in Eq. (17).

3. Methodology

3.1. Cases and Data

This study illustrates the comparative research with 2 simplified design cases. Case 1 is based on a design project for analyzing relationships between customer requirements and product configurations of mouse device, which is supported by a computer peripheral products company. There are 25 kinds of mouse devices provided by company that determined by 17 attributes which are selected by designers and marketers. Those attributes include price, net weight, dpi, color, working mode, etc. 13809 instances are confirmed the validation based on transaction data from May, 2015 to September, 2015 in an E-business transaction platform. Thus customer requirements on each attribute are determined by their choices on specific products.

Case 2 is a new product development (NPD) project for air conditioner that supported by a home appliance company. At first step, 460 customers are invited to raise their requirements on air conditions, and express their evaluations on the products. Then, show them with 12 conceptual products, and let them select the one that satisfy their requirements at best. Totally, 154 different attributes are determined based on customer evaluation criteria. At last, their selection on products and the corresponding requirements on attributes are processed for the purpose to map customer requirements to product configurations.

Table 1. Basic Data of the Two Cases

Case	Number of instances	Number of attributes	Number of classes	Product
Case 1	13809	17	25	Mouse device
Case 2	460	154	12	Air conditioner

3.2. Analysis

This study applies 10-fold cross-validation to conduct the comparison. For each case, nine tenths of the data are selected randomly as the training data, and the rest one tenth are acted as the testing role. All the combinations of dimensionality reduction methods and NBC are processed for ten times.

Case 1 chose randomly 12428 instances for training, and 1381 instances for testing in each round. Case 2 chose randomly 414 instances for training, and 46 instances for testing in each round.

3.3. Measures

This study applies 2 criteria to conduct measurements, they are accuracy and F-measure. These criteria are applied to measure the outcomes of the combination analysis of reduction methods and NBC.

NBC makes the idealized assumption that all attributes are conditionally independent given the class. Thus, the relationships of data after dimensionality reducing influence the performance of NBC in a manner. Let π denotes the overall accuracy of all classes, m denote the number of class, TC_i denotes the number of classifications that correctly assigned to class i, FC_i denotes the number of classifications that not belong to class i, then, the overall accuracy rate is given by:

$$\pi = \frac{\sum_{i=1}^{m} TC_i}{\sum_{i=1}^{m}\left(TC_i + FC_i\right)} \tag{37}$$

Accuracy omits the false negative rate when measures the performance of classifier [1]. F-measure is applied in this study to overcome the issue. However, F-measure is originally defined for binary classification case [18]. This study adopts macro-average F-measure proposed by [19] for multiple classification cases. F-measure ranges from 0 to 1, with higher value means better classification quality.

Let ρ_i denotes the recall rate of class i, FI_i denotes number of individuals which are not assigned to class i while they actually belong to. Then, the recall rate is given by:

$$\rho_i = \frac{TC_i}{TC_i + FI_i} \tag{38}$$

Suppose π_i is the accuracy of classification for class i, then the macro-average F-measure F_{macro} is calculated as:

$$F_{macro} = \frac{\sum_{i=1}^{m}\left(\frac{2\rho_i\pi_i}{\rho_i + \pi_i}\right)}{m} \tag{39}$$

4. Results

Table 2 and Table 3 display the average values of accuracy and F-measure for each scenario of combination of NBC with dimensionality reduction methods. In the Figure 1, we display the performance of PCA+NBC and LLE+NBC in every round for the two

cases, due to PCA and LLE performance better than other linear method and other nonlinear methods respectively. The findings are summarized in the following.

Table 2. Results of All Scenarios in Case 1 with Average Values

Measurement	PCA +NBC	ICA +NBC	KPCA +NBC	MDS +NBC	Isomap +NBC	LLE +NBC	LE +NBC
Accuracy (%)	86.89	84.90	60.83	60.17	69.98	76.39	74.08
F-measure (%)	89.12	85.44	61.35	60.83	68.74	77.27	75.21

Table 3. Results of All Scenarios in Case 2 with Average Values

Measurement	PCA +NBC	ICA +NBC	KPCA +NBC	MDS +NBC	Isomap +NBC	LLE +NBC	LE +NBC
Accuracy (%)	71.52	64.70	59.17	57.35	65.52	73.69	71.83
F-measure (%)	63.26	60.28	52.59	51.52	62.81	67.17	65.42

Firstly, all the reduction methods perform poorly in the case with relatively small samples and large attributes. The reason can be explained that dimension in such cases needed to be reduced to a very low dimension, which may result in loss of more information [12, 20]. Moreover, linear methods suffer more decrease of performance from case 1 to case 2, which indicates that linear methods are more sensitive to the loss of information.

Secondly, linear methods (PCA and ICA) perform better than the rest five nonlinear methods when combined with NBC in case 1, where dataset is relatively rich for each class. Some research also agrees that nonlinear dimensionality reduction methods perform poorly in realistic cases with various natural datasets [21]. However, in case 2, where there is larger amount of attributes and less sample in each class, two nonlinear methods, LLE and LE outperform other methods, including linear methods, in addition, Isomap performs better than ICA in this case. Reasons are various: (*a*) The two nonlinear methods, KPCA and MDS are incapable of modelling complex nonlinear manifolds [10]. (*b*) Though Isomap, LLE and LE are susceptible to the curse of dimensionality, over-fitting and the presence of outliers [10], they are based on construction of neighborhood graphs, and good in attaining information from local data distribution.

Thirdly, both in case 1 and case 2, the local linear methods (LLE and LE) with NBC outperform than global nonlinear method (Isomap). The reason is that the pairwise geodesic distances can be heavily influenced when Isomap suffers from short-circuiting referring to erroneous connections in neighborhood graph [22], while, LLE and LE just retain local properties of dataset, which makes them less sensitive to short-circuiting. Thus, Isomap is poorly performance for mapping high dimensional data to low dimensional data space than LLE and LE.
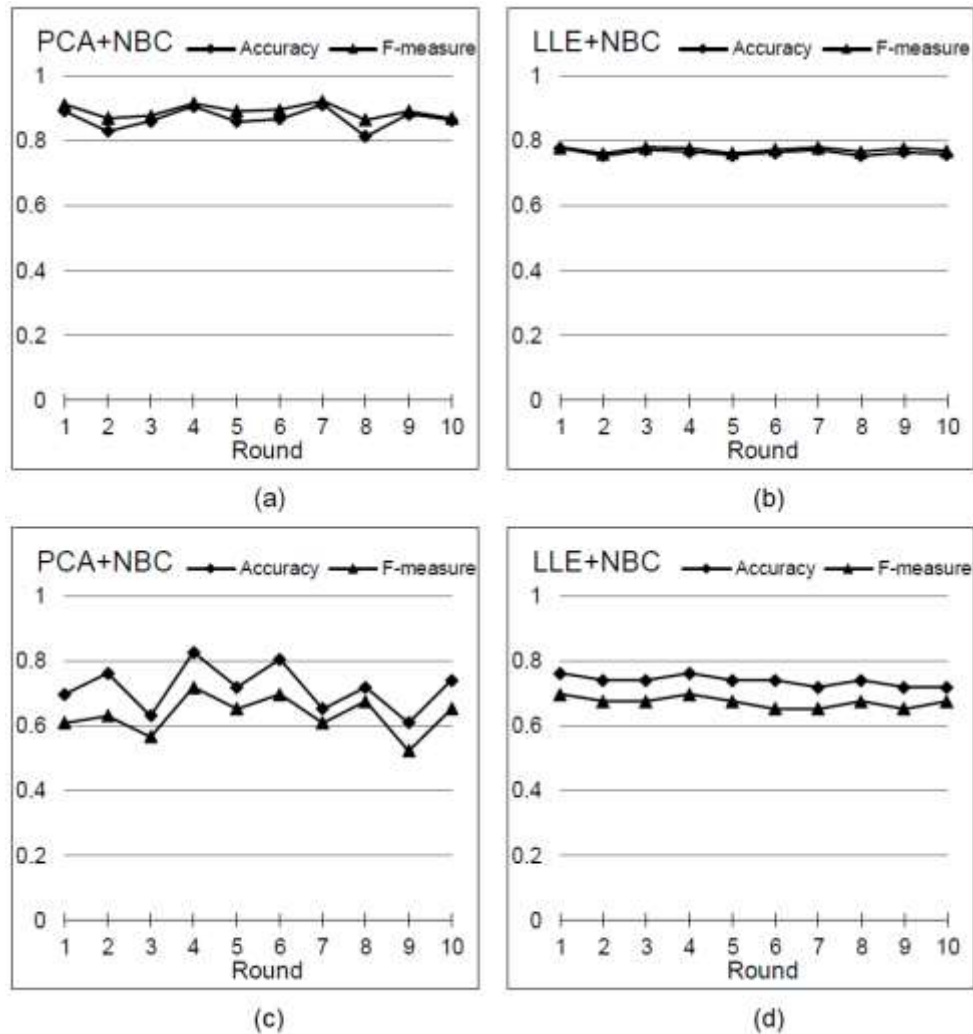
Figure 1. Performance of PCA and LLE with NBC in Each Round in Two Cases; (a) PCA+NBC in Case 1; (b) LLE+NBC in Case 1; (c) PCA+NBC in Case 2; (d) LLE+NBC in Case 2.

5. Conclusion

This paper presents a comparative study for evaluating 7 preprocessing dimensionality reduction methods, namely PCA, ICA, KPCA, MDS, Isomap, LLE and LE, for integrating with NBC to map customer requirements to product configurations. The comparison includes 2 realistic design cases and 2 performance measures.

At first, this study demonstrates how to use NBC to map customers' uncertain requirements to product configurations. In the next, 7 popular dimensionality reduction methods were introduced. Then these 7 methods are integrated as preprocess procedures for NBC to map customer requirements to product configurations in 2 design projects. The performances of integrative combinations are measured by the accuracy and F-measure. The results in this study show that the loss of information has great impact on the performance of dimensionality reduction methods, and linear methods are more sensitive to the loss of information. Also, the results indicate that several nonlinear methods, e.g., Isomap, LLE and LE, are good in handling the loss of information than other four methods. Moreover, local linear methods are suggested compared with global nonlinear methods.

Mapping customer requirements to product configuration is never an easy task in product design. The task is tougher in the issues of complex product design, for the much larger amount of attributes and much less amount of valid samples. In such cases, the performance of different dimensionality reduction method for integrating with NBC to solve the mapping task should be revisited to check.

Acknowledgments

References

[1] Y. Wang and M. M. Tseng, "A Naïve Bayes Approach to Map Customer Requirements to Product Variants", Journal of Intelligent Manufacturing, vol. 26, no. 3, (2015), pp. 501-509.

[2] B. Schwartz and K. Kliban, Editors, "The Paradox of Choice: Why More Is Less", ECCO, New York, (2005).

[3] Y. Wang and M. M. Tseng, "Integrating Comprehensive Customer Requirements into Product Design", CIRP Annals - Manufacturing Technology, vol. 60, no. 1, (2011), pp. 175-178.

[4] N. Friedman, G. Dan, and M. Goldszmidt, "Bayesian Network Classifiers", Machine Learning, vol. 29, no. 2-3, (1997), pp. 131-163.

[5] D. Yankelovich and D. Meer, "Rediscovering Market Segmentation", Harvard Business Review, vol. 84, no. 2, (2006), pp. 122-131.

[6] C. C. Yang, "An Analytical Methodology for Identifying the Latent Needs of Customers". Total Quality Management & Business Excellence, vol. 24, no. 11-12, (2013), pp. 1332-1346.

[7] E. Keogh and A. Mueen, "Encyclopedia of Machine Learning", Edited C. Sammut and G. I. Webb, Springer, New York, (2011), pp. 257-258.

[8] D. Mladenić, "Subspace, Latent Structure and Feature Selection", Edited C. Saunders, M. Grobelnik, S. Gunn and J. Shawe-Taylor, Springer-Verlag, Berlin, (2006), pp. 84-102.

[9] I. T. Jolliffe, Editor, "Principal Component Analysis", Springer, New York, (2002).

[10] L. J. P. van der Maaten, E. O. Postma and H. J. van den Herik, "Dimensionality Reduction: A Comparative Review", Centre for Creative Computing Tilburg, Tilburg, Netherlands, (2009).

[11] J. Stone, Editor, "Independent Component Analysis: A Tutorial Introduction", MIT Press, Cambridge, MA, (2004).

[12] A. B. Musa, "A Comparison of $\ell 1$-Regularizion, PCA, KPCA and ICA for Dimensionality Reduction in Logistic Regression", International Journal of Machine Learning and Cybernetics, vol. 5, no. 6, (2014), pp. 1-13.

[13] E. E. O. Ishida and R. S. D. Souza, "Kernel PCA for Type la Supernovae Photometric Classification", Monthly Notices of the Royal Astronomical Society, vol. 430, no. 1, (2012), pp. 509-532.

[14] Y. W. Xiao and X. H. Zhang, "Novel Nonlinear Process Monitoring and Fault Diagnosis Method Based on KPCA–ICA and MSVMs", Journal of Control, Automation and Electrical Systems, vol. 27, no. 3, (2016), pp. 289-299.

[15] C. Orsenigo, "An Improved Set Covering Problem for Isomap Supervised Landmark Selection", Pattern Recognition Letters, vol. 49, no. 49, (2014), pp. 131-137.

[16] Q. Li and H. Ji, "Multimodality Image Registration Using Local Linear Embedding and Hybrid Entropy", Neurocomputing, vol. 111, no. 6, (2013), pp. 34-42.

[17] A. M. Quispe, C. Petitjean and L. Heutte, "Extreme Learning Machine for Out-of-sample Extension in Laplacian Eigenmaps", Pattern Recognition Letters, vol. 74, (2016), pp. 68-73.

[18] G. Hripcsak and A. S. Rothschild, "Agreement, the F-measure, and Reliability in Information Retrieval". Journal of the American Medical Informatics Association, vol. 12, no. 3, (2005), pp. 296-298.

[19] A. Özgür, L. Özgür and L. Güngör, "Text Categorization with Class-based and Corpus-based keyword selection", Proceedings of the 20th International Symposium of Computer and Information Science, Istanbul, Turkey, (2005), October 26-28.

[20] L. Fan and K. L. Poh, "Computational and Ambient Intelligence", Edited F. Sandoval, A. Prieto, J. Cabestany and M. Grañ, Springer-Verlag, Berlin, (2007), pp. 16-22

[21] I. S. Lim, P. D. H. Ciechomski, S. Sarni and D. Thalmann, "Planar Arrangement of High-dimensional Biomedical Data Sets by Isomap Coordinates", Proceedings of the IEEE Symposium on Computer-Based Medical Systems, New York, USA, (2003), June 26-27.

[22] J. A. Lee and M. Verleysen, "Nonlinear Dimensionality Reduction of Data Manifolds with Essential Loops", Neurocomputing, vol. 67, no. 1, (2005), pp. 29-53.

Authors

Yao Jiao, He received his bachelor degree in Industrial engineering (2012) from Chongqing University. Now he is a doctoral candidate of Chongqing University. His current scientific interests include product design, customer requirements elicitation, intelligent system, and data mining.

Yu Yang, He received his master degree in Mechanical Engineering (1995) and PhD in Mechanical Engineering (1999) from Chongqing University. Now he is a Professor of Chongqing University (Since 2003). His current research interests include product design, intelligent system, data mining and mechanical design.