Expert Systems with Applications 39 (2012) 11634–11641

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Kernel based regression and genetic algorithms for estimating cutting conditions of surface roughness in end milling machining process

Antoni Wibowo *, Mohammad Ishak Desa

Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

ARTICLE INFO

Keywords:
Surface roughness
Milling
Kernel
Kernel principal component analysis
Nonlinear regression
Multicollinearity
Genetic algorithms

ABSTRACT

We observe a surface roughness in end milling machining process which is influenced by machine parameters, namely radial rake angle, speed and feed rate cutting condition. In this machining, we need to minimize and to obtain as low as possible the surface roughness by determining the optimum values of the three parameters. In previous works, some researchers used a response surface methodology (RSM) and a soft-computing approach, which was based on ordinary linear regression and genetic algorithms (GAs), to estimate the minimum surface roughness and its corresponding values of the parameters. However, the construction of the ordinary regression models was conducted without considering the existence of multicollinearity which can lead to inappropriate prediction. Beside that it is known the relation between the surface roughness and the three parameters is nonlinear, which implies that a linear regression model can be inappropriate model to approximate it. In this paper, we present a technique developed using hybridization of kernel principal component analysis (KPCA) based nonlinear regression and GAs to estimate the optimum values of the three parameters such that the estimated surface roughness is as low as possible. We use KPCA based regression to construct a nonlinear regression and to avoid the effect of multicollinearity in its prediction model. We show that the proposed technique gives more accurate prediction model than the ordinary linear regression's approach. Comparing with the experiment data and RSM, our technique reduces the minimum surface roughness by about 45.3% and 54.2%, respectively.

1. Introduction

Surface roughness is one of the most common performance measurements in a machining process and an effective parameter to measure the quality of machined surface. A machined surface has the higher quality than another machined surface when it has lower surface roughness than the other one. In machining, we need to have as low as possible of the surface roughness by setting the suitable values of the influenced parameters. Hence, improvement in the quality could be indicated by referring to the performance of the influenced parameters in its surface roughness. The conventional optimization approach can optimize machining problem by using Taguchi technique, factorial technique and response surface methodology (RSM) technique (Mukherjee & Ray, 2006). The new trends of optimization techniques are soft-computing based optimization, including genetics algorithms (GAs), Simulated Annealing (SA), Tabu Search, Particle Swarm Optimization (PSO) and Neural Networks (NN), which can be used to estimate the highest quality of machined surface (Mukherjee & Ray, 2006; Zain, Haron, & Syarif, 2010).

Recently, Zain et al. (2010) proposed an optimization technique based on ordinary linear regression and GAs to estimate the quality of surface roughness. They used the experiment data conducted by Mohruni (2008) in which the surface roughness is influenced by radial rake angle, speed and feed rate cutting condition of the machine used. Their technique was performed by developing linear regression models using the Mohruni's data, and followed by model selection among the regression models. Then, they developed a linear programming problem in which the objective function was one of their regression models and the constraints were developed based on the range of radial rake angle, speed and feed rate cutting condition data.

It is known that one important issue in ordinary linear regression is the existence of multicollinearity in its regressor matrix. We say that a multicollinearity exists on a matrix regressors, say $\mathbf{Z}$, if some eigenvalues of $\mathbf{Z}^T\mathbf{Z}$ are close to zero. In practice, when the ratio between a eigenvalue of $\mathbf{Z}^T\mathbf{Z}$ and the largest eigenvalue of $\mathbf{Z}^T\mathbf{Z}$ of is less than 1/1000 then a severe multicollinearity exists in $\mathbf{Z}$ (Montgomery, Peck, & Vining, 2006). When we use ordinary linear regression in the Mohruni's data, we can find that the ratios of eigenvalues and the largest eigenvalue of its regressor matrix are 1.329×10^{-8}, 3.036×10^{-7} and 3.249×10^{-4}, respectively, which indicate the existence of a severe multicollinearity in this regressor

* Corresponding author.
E-mail addresses: antoni@utm.my (A. Wibowo), mishak@utm.my (M.I. Desa).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.004

matrix. When a severe multicollinearity exists in a regressors matrix, it can give the negative effects to the corresponding regression model such as its prediction model can be inappropriate to be used (Montgomery et al., 2006).

We also noticed that the ideal surface roughness model is given by (Zain et al., 2010)

$$R_a = cv^k f^l \gamma^m \varepsilon \tag{1}$$

where R_a is the experiment (measured) surface roughness (μm), v is the cutting speed (m/min), f is the feed rate (mm/tooth), γ is the radial rake angle (°), ε is the experimental error and c, m, k and l are the model parameters to be estimated, respectively. It is obvious that model (1) is, itself, nonlinear which implies that a linear regression model can be inappropriate model to approximate it.

In this paper, we present a technique using hybridization of *kernel principal component analysis* (*KPCA*) based regression and GAs to estimate the suitable values of the three parameters such that the estimated surface roughness is as low as possible. We use KPCA based regression, usually referred as *kernel principal component regression* (*KPCR*), to overcome the effect of linearity and multicollinearity in regression. KPCA has been used for nonlinear systems by mapping an original input space into a higher-dimensional feature space (Cho, Lee, Choi, Lee, & Lee, 2005; Lu, Zhang, Zhang, & Zhang, 2007; Scholkopf, Smola, & Muller, 1998; Scholkopf & Smola, 2002). KPCA becomes an attractive algorithm because it does not involve nonlinear optimization, it is as simple as the *principal component analysis* (*PCA*) and it does not need to specify the number of principal components prior to modeling compared to other nonlinear methods. KPCA based regression was studied by Rosipal, Girolami, Trejo, and Cichoki (2001), Rosipal and Trejo (2002), Jade, Srikanth, Kulkari, Jog, and Priya (2003), Hoegaerts, Suykens, Vandewalle, and Moor (2005), Wibowo (2008) and Wibowo and Desa (2011a, 2011b). which can be used to perform a nonlinear prediction and to dispose the effects of multicollinearity in regression model. Afterward, we use the *Cross Validation* (*CV*) method to select the 'best' regression model, and followed by developing nonlinear programming problems in which GAs are performed to estimate the surface roughness and its optimum solutions.

The rest of the manuscript is organized as follows. In Section 2, we review theories and methods of linear regression, KPCA, KPCA based regression including its algorithm, and followed by GAs. In Section 3, we present our case study using Mohruni's data, model selection and constructing nonlinear programming problem of the surface roughness. We end this section by finding the minimum of the surface roughness and its corresponding variables. Finally, conclusions are given in Section 4.

2. Theories and methods

2.1. Linear regression

Suppose that we have the data $(y_i \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{iN})^T \in \mathbb{R}^{p+1}$ where $i = 1,2,\ldots,N$; N is the number of data and $\mathbb{R}$ is the set of real numbers. Then, we define $\mathbf{y} = (y_1 \quad y_2 \quad \cdots \quad y_N)^T \in \mathbb{R}^N$, $\mathbf{X} = \left(\mathbf{1}_N \quad \widetilde{\mathbf{X}}\right) \in \mathbb{R}^{N\times(p+1)}$ where $\widetilde{\mathbf{X}} = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_N) \in \mathbb{R}^{N\times p}$, 1_N is $N \times 1$ vector with all elements equal to one, $\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{iN})^T \in \mathbb{R}^p$. The ordinary multiple linear regression model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{2}$$

where $\boldsymbol{\beta} = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_p)^T \in \mathbb{R}^{p+1}$ is a vector of regression coefficients and $\mathbf{e} = (e_1 \quad e_2 \quad \cdots \quad e_N)^T \in \mathbb{R}^N$ is vector of residuals. The matrix $\mathbf{X}$ is called the matrix regressors.

When we use the *ordinary least squares* (*OLS*) method to find the estimate of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}$, then the estimate is found by minimizing

$$\sum_{i=1}^{N} e_i^2 \tag{3}$$

where $e_i = y_i - \dot{\mathbf{x}}_i^T\boldsymbol{\beta}$ and $\dot{\mathbf{x}}_i^T = (1 \quad \mathbf{x}_i^T)$. The solution can be found by solving the following linear equation

$$\mathbf{X^T X}\hat{\boldsymbol{\beta}} = \mathbf{X^T y} \tag{4}$$

The prediction of linear regression is given by

$$g(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i x_i \tag{5}$$

where g is a function from $\mathbb{R}^N$ into $\mathbb{R}$ and $\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_p)^T$. When multicollinearity exists on $\boldsymbol{X}$ then the prediction $g(\mathbf{x})$ can be inappropriate to be used (Montgomery et al., 2006) and called it the *effect of multicollinearity*.

2.2. Regression based on kernel principal component analysis (KPCA)

2.2.1. KPCA

Assume that we have a function $\psi : \mathbb{R}^p \rightarrow \mathbf{F}$, where $\mathbf{F}$ is the feature space which is a Euclidean space with dimension p_F ($p_F \geqslant p$). Then, we define the matrices $\mathbf{C} = (1/N)\mathbf{\Psi^T\Psi} \in \mathbb{R}^{p_F\times p_F}$ and $\mathbf{K} = \mathbf{\Psi\Psi^T} \in \mathbb{R}^{N\times N}$ where $\mathbf{\Psi} = (\psi(\mathbf{x}_1) \quad \psi(\mathbf{x}_2) \quad \cdots \quad \psi(\mathbf{x}_N))^T \in \mathbb{R}^{N\times p_F}$ and assume that $\sum_{i=1}^{N}\psi(\mathbf{x}_i) = \mathbf{0}$. The relation of eigenvalues and eigenvectors of the matrices $\mathbf{C}$ and $\mathbf{K}$ were studied by Scholkopf et al. (1998) and Wibowo (2008).

Let $\hat{p}_F$ be the rank of $\mathbf{\Psi}$ where $\hat{p}_F \leqslant \min(N, p_F)$ which implies that both rank ($\mathbf{K}$) and rank ($\mathbf{\Psi^T\Psi}$) are equal to $\hat{p}_F$. It is evident that the eigenvalues of $\mathbf{K}$ are nonnegative real numbers since the matrix $\mathbf{K}$ is symmetric and positive semidefinite (see Anton (2000) for the detailed discussion of symmetric and positive semidefinite matrix). Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{\hat{r}} \geqslant \lambda_{\hat{r}+1} \geqslant \cdots \geqslant \lambda_{\hat{p}_F} > \lambda_{\hat{p}_F+1} = \cdots = \lambda_N = 0$ be the eigenvalues of $\mathbf{K}$ and $\mathbf{B} = (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_N)$ be the matrix of the corresponding normalized eigenvectors $\mathbf{b}_s (s = 1,2,\ldots,N)$ of $\mathbf{K}$. Then, let $\boldsymbol{\alpha}_l = \mathbf{b}_l/\sqrt{\lambda_l}$ and $\mathbf{a}_l = \mathbf{\Psi^T}\boldsymbol{\alpha}_l$ for $l = 1,2,\ldots,\hat{p}_F$. The eigenvectors $\mathbf{a}_l$, however, cannot be found explicitly since we do not know $\mathbf{\Psi^T\Psi}$ explicitly. However, we can obtain the principal component of $\psi(\mathbf{x})$ corresponding to nonzero eigenvalues of $\mathbf{\Psi^T\Psi}$ by using a *kernel trick*. The lth principal component of $\psi(\mathbf{x})$ ($l = 1,2,\ldots,\hat{p}_F$) is given as follows:

$$\psi(\mathbf{x})^T\mathbf{a}_l = \sum_{i=1}^{N}\alpha_{li}\psi(\mathbf{x})^T\psi(\mathbf{x}_i) \tag{6}$$

where α_{li} is the ith element of $\boldsymbol{\alpha}_l$. According to Mercer Theorem (Scholkopf et al., 1998; Scholkopf & Smola, 2002), if we choose a continuous, symmetric and positive semidefinite kernel $\kappa : \mathbb{R}^P \times \mathbb{R}^P \rightarrow \mathbb{R}$ then there exists $\phi : \mathbb{R}^P \rightarrow \mathbf{F}$ such that $\kappa(\mathbf{x}_i,\mathbf{x}_j) = \varphi(\mathbf{x}_i)^T\varphi(\mathbf{x}_j)$. Instead of choosing ψ explicitly, we choose a kernel κ and employ the corresponding function φ as ψ. Let $K_{ij} = \kappa(\mathbf{x}_i,\mathbf{x}_j)$ then $\mathbf{K}$ and $\boldsymbol{\alpha}_l (l = 1,2,\ldots,\hat{p}_F)$ are explicitly known now. Therefore, Eq. (6) is also explicitly known and can be written as

$$\psi(\mathbf{x})^T\mathbf{a}_l = \sum_{i=}^{N}\alpha_{li}\kappa(\mathbf{x},\mathbf{x}_i) \tag{7}$$

2.2.2. KPCA Based Regression

The centered multiple linear regression in the feature space is given by

$$\mathbf{y}_0 = \mathbf{\Psi}\boldsymbol{\gamma} + \tilde{\mathbf{e}} \tag{8}$$

where $\boldsymbol{\gamma} = (\gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_{P_F})^T$ is a vector of regression coefficients in the feature space, $\tilde{\mathbf{e}}$ is a vector of random errors and