Sensors & Transducers, Vol. 173, Issue 6, June 2014, pp. 224-229

Sensors & Transducers

Soft Sensor Modeling Based on Multiple Gaussian Process Regression and Fuzzy C-mean Clustering

Xianglin ZHU, Xiaofu JI
School of Electrical and Information Engineering, Jiangsu University,
212013, P. R. China

Received: 5 June 2014 /Accepted: 27 June 2014 /Published: 30 June 2014

Abstract: In order to overcome the difficulties of online measurement of some crucial biochemical variables in fermentation processes, a new soft sensor modeling method is presented based on the Gaussian process regression and fuzzy C-mean clustering. With the consideration that the typical fermentation process can be distributed into 4 phases including lag phase, exponential growth phase, stable phase and dead phase, the training samples are classified into 4 subcategories by using fuzzy C-mean clustering algorithm. For each sub-category, the samples are trained using the Gaussian process regression and the corresponding soft-sensing sub-model is established respectively. For a new sample, the membership between this sample and sub-models are computed based on the Euclidean distance, and then the prediction output of soft sensor is obtained using the weighting sum. Taking the Lysine fermentation as example, the simulation and experiment are carried out and the corresponding results show that the presented method achieves better fitting and generalization ability than radial basis function neutral network and single Gaussian process regression model.

Keywords: Gaussian process regression, Fuzzy C-mean clustering, L-lysine, Soft sensor.

1. Introduction

The fermentation process is a highly nonlinear and uncertain system since some complex biological, physical and thermodynamic reactions are involved simultaneously. In this process, some important variables reflecting the fermentation quality, such as cell concentration, substrate concentration and product concentration are difficult to detect online. The traditional method to obtain these variables is online sample and offline analysis that is characterized by time-consumption and then it cannot realize online detect. Meanwhile, the online sample method may introduce bacterial contamination that tremendously degrades the fermentation quality.

A new method called soft sensing can efficiently resolve this problem [1, 2]. The idea of soft sensing originated from the inferential control theory where primary variables difficult to detect online is numerically estimated by some assistant variables easy to detect through measuring instruments. The commonly used soft sensing model can be classified into 3 categories: mechanism model, data-driven model and mixed model [3]. Among these methods, the data-driven method builds the soft sensing model based on lots of process data instead of exact process model, which is very suitable to fermentation process with unknown reaction mechanism. The frequently applied data-driven soft-sensing methods include artificial neural networks (ANNs), support vector machines (SVMs), and Gaussian process regression respectively. The Gaussian process repression applies the prior Gaussian distribution to the nonparametric regression function space, and builds the statistical

soft-sensing model by calculating the posterior distribution of the prediction function. Compared with artificial neural networks and support vector machines, the Gaussian process regression method has some outstanding advantages, such as less parameters to be optimized, fast convergence speed, high model precision and good generalization ability [5, 7-9].

A soft sensing method is given for fermentation processes in this paper based on the Gaussian process regression and fuzzy C-mean clustering. With the consideration that the typical fermentation process can be distributed into 4 phases, that is, lag phase, exponential growth phase, stable phase and dead phase, the process samples is classified into 4 sub-categories using the fuzzy C-mean clustering method. For each subcategory, the corresponding soft sensor is built using the Gaussian process regression method. For a new sample, the membership value of every Gaussian process model is first calculated by computing the distance between this new sample and every clustering center, and then the prediction output can be computed. To illustrate the effectiveness of the proposed method, the presented method is used to estimate the cell concentration in the Lysine fermentation process. The simulation shows that the presented soft sensing method achieves much better fitting performance and generalization ability than radial basis function neural networks and single Gaussian process model.

2. Gaussian Process and Regression Algorithm

2.1. Fundamental Theory of Gaussian Process

The Gaussian process is a novel machine learning algorithm developed recently on the basis of the Gaussian stochastic process and Bayesian learning theory. It describes such a stochastic process that the arbitrary finite set follows the Gaussian distribution [5]. For any integer $n \geq 1$ and stochastic variable X, the corresponding joint probability distribution of process state $f(X)$ at time t follows n-dimensional Gaussian one. The whole statistics characteristic is governed by sample mean $m(t)$ and covariance function $k(t,t')$, that is,

$$f(t) \sim \mathrm{GP}(m(t), k(t,t')), \quad (1)$$

2.2. Gaussian Process Regression

For a given training sample set $D = \{(x_i, y_i) \mid i = 1,2,\ldots,n\}$, where $x_i \in \mathcal{R}^d$ is the d-dimensional input variable, $y_i \in \mathcal{R}$ is the corresponding output variable. In the context of soft sensor design, x_i denotes the assistant variable and y_i denotes the primary variable. For notational simplicity, we use X to denote $d \times n$ input matrix, y to denote output vector. Then, the training sample set can be expressed as $D = (X, y)$. For a new sample x^*, the corresponding model output y^* can be obtained through the prior knowledge of $D = (X, y)$.

Assume that the observed target value y is contaminated by noise and that the difference between them is ε, that is,

$$y = t + \varepsilon, \quad (2)$$

where ε is the independent random variable following the Gaussian distribution with mean 0 and variance σ^2, that is,

$$\varepsilon \sim \mathcal{N}(0, \sigma^2), \quad (1)$$

The prior distribution of the observed target value y is

$$y \sim \mathcal{N}(0, K + \sigma^2 I), \quad (1)$$

where $K = K(X,X) = (k_{ij})_{n \times n}$ is the symmetric positive-definite covariance matrix, and k_{ij} is the scalar measuring the relationship between x_i and x_j.

The joint Gaussian distribution of y^* and y can be given as

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \left(0, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X,x^*) \\ K^T(X,x^*) & K(x^*,k^*) \end{bmatrix}\right), \quad (5)$$

where $K(X,x^*)$ is the joint covariance matrix of X and x^*, $K(x^*,x^*)$ is the auto-covariance matrix of x^*.

In the Gaussian process regression, the covariance function must satisfy the condition that a semi-positive definite covariance function should exist for an arbitrary sample. The frequently used covariance function is

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq}, \quad (6)$$

where hyper-parameters l, σ_f and σ_n greatly affect the generalization ability of the Gaussian process model. The frequently used method to obtain the optimal hyper-parameters is the maximum likelihood one, that is, to calculate the partial derivative of log-likelihood function with respect to hyper-parameters

and search the optimal hyper-parameters using some search algorithm such as the conjugate gradient algorithm and partial swarm optimization algorithm. The log-likelihood function is given as

$$\begin{aligned} L &= \ln p(y \mid X) \\ &= -\frac{1}{2} y^T (K + \sigma^2 I)^{-1} y - \frac{1}{2} \ln \left| K + \sigma_n^2 I \right| - \frac{n}{2} \ln 2\pi \end{aligned}, \quad (7)$$

For a new input x^*, the corresponding prediction output y^* can be obtained using the Bayesian theory,

$$p(y^* \mid x^*, X, y) \sim N(\hat{y}(x^*, \sigma(x^*))) \ , \quad (8)$$

where the mean $\hat{y}(x^*)$ and variance $\hat{\sigma}(x^*)$ are given by

$$\hat{y}(x^*) = k^T(x^*)(K + \sigma_n^2 I)^{-1} y \, , \quad (9)$$

$$\hat{\sigma}(x^*) = k(x^*, x^*) - k^T(x^*)(K + \sigma_n^2 I)^{-1} k(x^*) \, , \quad (10)$$

3. Fuzzy C-mean Clustering Algorithm

The fuzzy C-mean clustering algorithm is presented by Dunn, and developed by Bezdek in [16], whose basic principle is to classify unlabeled samples by minimizing the norm- and prototype-based objective function. For a given sample set $X = \{x_1, x_2, \ldots, x_n\} \subset \mathcal{R}^s$, where s is the dimension of sample space, n is the number of samples. Let $c > 1$ be the number of clustering. Then, the objective function of fuzzy C-mean clustering algorithm can be described as

$$\min f_{\text{fcm}}(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \, , \quad (11)$$

with constraints

$$\begin{cases} \sum_{i=1}^{c} u_{ij} = 1, 1 \le j \le n \\ \sum_{j=1}^{n} u_{ij} > 0, 1 \le i \le c \\ u_{ij} \ge 0, 1 \le i \le c, 1 \le j \le n \end{cases} , \quad (12)$$

where $m \ge 1$ is the fuzzy coefficient, $U = (u_{ij})_{c \times n}$ is the fuzzy dividing matrix, u_{ij} is the fuzzy membership value of the sample x_j to the ith class, $V = [v_1, v_1, \ldots, v_c]$ is a matrix composed of clustering center vector, $d = \| x_j - v_i \|$ is the distance between the sample x_j and the clustering center v_i. For optimization problem (11), the following iterative equation can be obtained using KKT condition,

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}, i = 1, 2, \ldots, c \, , \quad (13)$$

Let $I_j = \{(i, j) \mid x_j = v_i, 1 \le i \le c\}$. If $I_j = \varnothing$, then

$$u_{ij} = \left(\sum_{r=}^{c} \left(\frac{d_{ij}}{d_{rj}} \right)^{\frac{2}{m-1}} \right)^{-1} \, , \quad (14)$$

Otherwise, u_{ij} must be a nonnegative real scalar satisfying the following condition,

$$\sum_{i=1}^{c} u_{ij} = 1, u_{ij} = 0, d_{ij} \ne 0 \, , \quad (15)$$

This iterative equation can be regarded as a map from a point to a set. From a computational point of view, the following iterative algorithm of membership function is applied,

$$u_{ij} = \begin{cases} u_{ij} = \left(\sum_{r=}^{c} \left(\frac{d_{ij}}{d_{rj}} \right)^{\frac{2}{m-1}} \right)^{-1}, \text{if } I_j = \varnothing \\ \frac{1}{|I_j|}, \text{if } I_j = \varnothing, i \in I_j \\ 0, \text{if } I_j = \varnothing, i \notin I_j \end{cases} , \quad (16)$$

In the fuzzy C-mean clustering algorithm, the clustering centers or membership matrix is first given, and (13) and (16) are repeated until the given end condition is satisfied. The detailed procedure is given as follows,

1) Set the clustering number c, the fuzzy coefficient m, the clustering center $V^{(0)}$, the convergence precision $\varepsilon > 0$, and iterative index $k = 1$;
2) Compute $U^{(k+1)}$ using (16);
3) Compute $V^{(k+1)}$ using (13). Set $k := k + 1$;
4) Repeat steps (2) and (3) until the following end condition is satisfied:

$$\| V^{(k)} - V^{(k+1)} \| \le \varepsilon, k \ge 1 \, , \quad (17)$$

4. Soft Sensing Model for Lysine Fermentation

In this section, we take the cell concentration of Lysine fermentation for example to illustrate the detailed modeling procedure of soft sensing model based on the Gaussian process regression and fuzzy C-mean clustering. With the consideration that this fermentation process can be distributed into 4 phases including lag phase, exponential growth phase, stable phase and dead phase, the preprocessed samples are classified into 4 subcategories $(c = 4)$ using fuzzy C-mean clustering algorithm and then the Gaussian process regression algorithm is used to build the soft sensing model for every category.

The detailed modeling procedure is described as follows:

1) Build the training sample set (x_i, y_i) based on process data, where $x_i \in \mathcal{R}^d$ is the d-dimensional input presenting some variables affecting the cell concentration and easy to measure online, $y_i \in \mathcal{R}$ is the cell concentration value.
2) Preprocess and normalize the sample data to [0,1] in order to improve the model precision. Denote the sample matrix as $\hat{U} = (\hat{u}_{iq})$.
3) Construct the fuzzy relation matrix $R = (r_{iq})$, where r_{iq} is the coefficient of similarity degree between the i th and q th sample. The distance is described as

$$r_{iq} = 1 - c\sqrt{\sum_{i=1}^{m}(x_{iq} - x_{jk})^2}\,, \tag{18}$$

4) Compute the fuzzy equivalence matrix R'. Use the square method to compute the $R^2, R^4, \ldots,$ until $R^{2k} = R^{2k-1}$ and then obtain $R' = R^{2k}$.
5) Obtain the original clustering using λ intercept method in matrix R'. Note here $c = 4$. Denote the average value of subset samples as clustering centers, and express as $V_1^{(0)}$, $V_2^{(0)}$, $V_3^{(0)}$ and $V_4^{(0)}$.
6) Compute the similarity degree between samples and clustering centers using (18).
7) Compute the membership matrix $U^{(0)}$ of the normalized sample matrix $\hat{U}$.
8) Repeat the fuzzy C-mean clustering until the given precision is satisfied.
9) Build the Gaussian process regression model for each subcategory.
10) For a new sample, compute the membership value of sample to each clustering center, compute prediction output y_i of each class, and obtain the cell concentration value as

$$y = \sum_{i=1}^{c} y_i\,, \tag{19}$$

5. Experiments and Simulation

The fermentation experiments are carried out based on WKT-30L fermentation equipment and the corresponding digital control system shown in Fig. 1.

Fig. 1. WKT-30L fermentation experiment system.

According to the process environment requirement, the pressure in fermentation process is controlled to be 0.11 MPa, temperature 31 °C, rotating speed of the mixing motor 220 r/min. According to early studies [12, 3, 15], the assistant variables are chosen as the variables of pH level (pH), dissolved oxygen (DO), air flux F and the primary variable is chosen as cell concentration y. In this way, the soft sensor model can be designed as

$$y = f(\mathrm{pH}, \mathrm{DO}, F)\,, \tag{20}$$

where $f(\cdot)$ denote the complex nonlinear relationship among these variables.

To obtain enough testing samples, WKT-30L experiment system samples pH, DO, and *F* every minute. At the same time, the fermentation liquid is sampled every 4 hours that gives exact cell concentration via 721-type spectrophotometer. One batch sample can be obtained in one fermentation period. Totally, 5 experiments are designed and thus 184 samples are obtained, among which, the former 138 samples are used as training data and the rest 36 samples are used as testing data. To obtain a better model performance, samples are normalized into the interval [0,1].

The 3D clustering result is shown in Fig. 2, where 4 classified subcategories are denoted by ellipses.

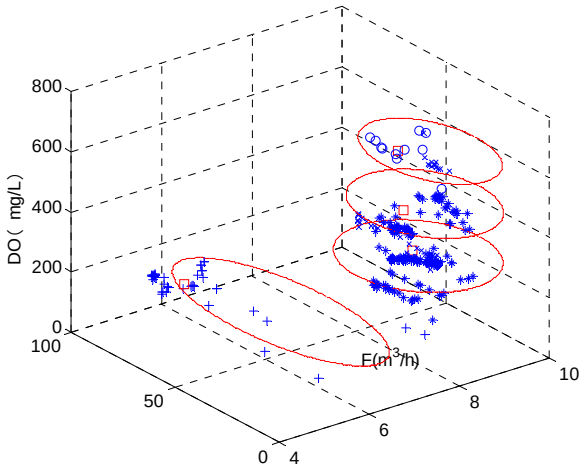

Fig. 2. 3D cluster results in training set.

The changing curve of objective function is shown in Fig. 3. After 44 iterations, the given precision of objective function is satisfied.

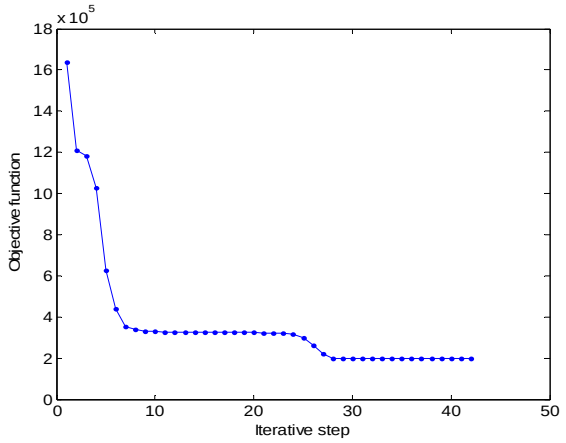

Fig. 3. Variation curves of the objective function value.

The real value and prediction value of cell concentration in Lysine fermentation is shown in Fig. 4.

To compare with others soft-sensing methods, the radial basis function neural networks and single Gaussian process regression methods are also applied to build the soft-sensing model as shown in Fig. 5. The relative error of 3 methods are shown in Fig. 6. It can be shown that the presented method has better fitting and generalization ability.

To further analyze the prediction ability of three models, the following average relative error (ARE), maximum absolute error (MaxE), and root mean squared error (MSE) are defined to evaluate the model performance,

$$\mathrm{ARE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i^*}{y_i}\right| \times 100\%, \tag{21}$$

$$\mathrm{MaxE} = \max(|\, y_i - y_i^* \,|), \tag{22}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2, \tag{23}$$

where y_i is the exact value, y_i^* is the prediction value.

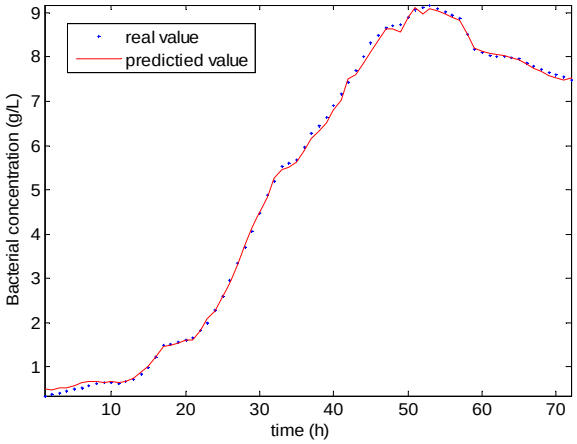

Fig. 4. Real value and predicted value using presented method.

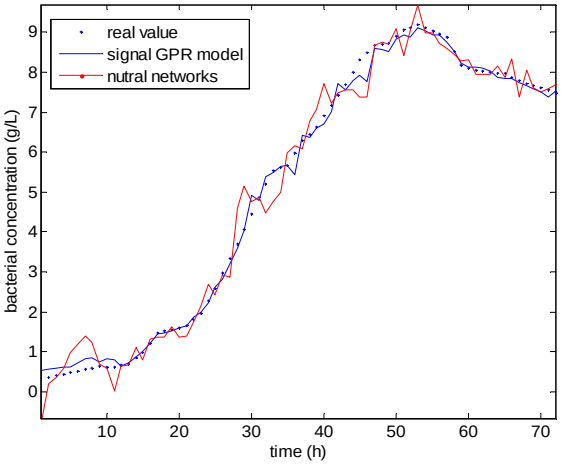

Fig. 5. Variation curves of the objective function value.

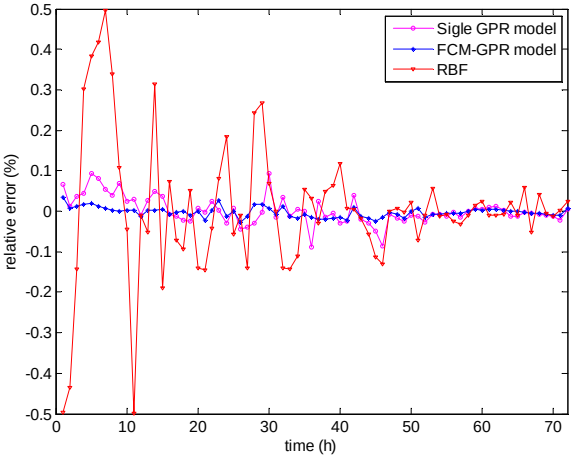

Fig. 6. Real value and predicted value using presented method.

The RMSE, ARE and MaxE of three soft-sensing models are given respectively in Table 1, which shows that the presented soft-sensing method in this paper achieves better generalization ability than radial basis function neural networks and single Gaussian process regression methods.

Table 1. MSE, ARE and MaxE of three soft-sensing models.

Method	ARE	RMSE	MAXE
Neural networks	0.1881	0.0491	1.1069
GPR	0.0767	0.0208	0.7218
Presented method	0.0365	0.0095	0.2049

6. Conclusion

A new soft-sensing method is presented for fermentation processes based on Gaussian process regression and fuzzy C-mean clustering. With the consideration that this fermentation process can be distributed into 4 phases including lag phase, exponential growth phase, stable phase and dead phase, the preprocessed samples are classified into 4 sub-categories using fuzzy C-mean clustering algorithm and then the Gaussian process regression algorithm is used to build the soft sensing model for every sub-category. This method efficiently overcomes some shortcomings of global soft-sensing method including long learning time, unmatched process characteristics, low prediction precision, and low adaptive ability. The simulation shows that the presented method achieves better prediction and generalization ability than radial basis function neural networks and single Gaussian process regression methods and that it can well realize the soft sensing of cell concentration in the Lysine fermentation process.

Acknowledgements

This work is supported by the natural science foundation of Jiangsu Province, China under grant BK2011465.

References

[1]. G. Liu, X. Jiang, C. Mei, Research on soft sensing method based on continuous hidden Markov model in fermentation process, *Control and Decision*, Vol. 26, Issue 11, 2011, pp. 1753-1760.

[2]. C. Brosillow, Inferential control of process, *American Institute of Chemical Engineers Journal*, Vol. 24, Issue 1, 1978, pp. 485-509.

[3]. Y. Sun, B. Wang, Soft sensor of Lysine fermentation based on fuzzy support vector machines, in *Proceedings of the Chinese Control Conference (CCC'08)*, Kunming China, 16-18 July, 2008, pp. 280-284.

[4]. L. Huang, Y. Sun, X. Ji, Y. Huang, Soft sensor modeling of fermentation process based on the combination of CPSO and LSSVM, *Chinese Journal of Scientific Instrument*, Vol. 32, Issue 1, 2011, pp. 2066–2070.

[5]. C. E. Rasmussen, C. K. T. Williams, Gaussian processes for machine learning, *MIT Press*, Massachusetts, 2006, pp.7–31.

[6]. M. Girolami, S. Rogers, Variational Bayesian multi-nomial probit regression with Gaussian process priors, *Neural Computation*, Vol. 18, Issue 1, 2006, pp. 1790-1817.

[7]. S. Brahim-Belhouari, A. Bermak, Gaussian process for non-stationary time series prediction, *Computational Statistic & Data Analysis*, Vol. 47, Issue 4, 2004, pp. 705-712.

[8]. B. Sun, H. Yao, T. Liu, Short-term wind speed forecasting based on Gaussian process regression model, *Proceedings of the Chinese Society of Electrical Engineering*, Vol. 32, Issue 29, 2012, pp. 104-109.

[9]. G. Su, Y. Zhang, L. Yan, Gaussian process model for forecasting bearing capacity of composite foundation with CFG pole, *Computer Engineering and Applications*, Vol. 47, Issue 4, 2011, pp. 236-238.

[10]. Y. Li, H. Yang, Ensemble modeling method based on Bagging algorithm and Gaussian process, *Journal of Southeast University (Natural Science Edition)*, Vol. 41, Issue 1, 2011, pp. 93-96.

[11]. F. Qu, Y. Hu, S. Ma, A convergence theorem of kernel based fuzzy C-means clustering algorithm, *Journal of Jilin University (Science Edition)*, Vol. 49, Issue 2, 2011, pp. 1079-1086.

[12]. L. Huang, Y. Sun, X. Ji, Y. Huang, Soft sensor of Lysine fermentation based on tPSO-BPNN, *Chinese Journal of Scientific Instrument*, Vol. 31, Issue 10, 2010, pp. 2317-2321.

[13]. Y. Huang, Y. Sun, B. Wang, Research of soft sensor based on fuzzy neural network inverse system for Lysine fermentation process, *Chinese Journal of Scientific Instrument*, Vol. 31, Issue 10, 2010, pp. 862-867.

[14]. B. Wang, Y. Sun, X. Ji, Soft-sensor modeling for Lysine fermentation processes based on PSO-SVM inversion, *CIESC Journal*, Vol. 63, Issue 9, 2012, pp. 3000-3007.

[15]. Y. Sun, M. Chen, X. Ji, Soft sensor of lysine fermentation biological parameters based on SVM, *Chinese Journal of Scientific Instrument*, Vol. 29, Issue 10, 2008, pp. 2067-2071.

[16]. J. C. Bezdek, Pattern recognition with fuzzy objective function algorithm, *Plenum Press*, New York, 1981.