An Intelligent Machine Monitoring System Using Gaussian Process Regression for Energy Prediction

Raunak Bhinge, Nishant Biswas and David Dornfeld
Mechanical Engineering
UC. Berkeley
Berkeley, CA, USA

Jinkyoo Park and Kincho H. Law
Civil and Environmental Engineering
Stanford University
Stanford, CA, USA

Moneer Helu and Sudarsan Rachuri
National Institute of Standards and Technology
Gaithersburg, MD, USA
Mileage for automobiles has greatly increased. We also understand the energy consumption pattern well.
What about manufacturing energy use?

Industrial: 22% of U.S. energy use
Can we predict how much energy the manufacturing machine will consume when machining the part?
From data to insight: Mapping control parameters to energy consumption

Part design → NC code → Target machine → Machined part

XML-based structured data with automatic acquisition

Input parameters for the machine
- Feed rate
- Spindle speed
- Depth of cut
- Cutting direction
- Cutting strategy
- Dimensions

Output measurements from the machine
- Energy consumption
- Part quality, e.g., surface roughness
- Machine conditions, e.g., tool wear
Procedure for constructing data-driven energy prediction model

\[ D = \{(x^i, y^i); i = 1, \ldots, m\} \]

\( x = \) control parameter
\( y = \) Response

Using training data set \( D \), learning algorithm finds the best function \( h(x) \) that is believed to accurately predict the output \( y \) for a given input \( x \)

**Predict total energy consumption**

\[ \sum_{i=1}^{m} h(x^i) \]
Data acquisition & simulator for this study

Hardware

- **Fanuc Controller**: Collect machine control parameters
- **System Insights**
  - **High Speed Power Meter (HSPM)**: Collect power time series

Data types

- **MTConnect data**
  - Direct data: Timestamp, Real power, Feed rate, Spindle speed, Block of code
  - Derived data: Duration, Energy, Avg Feed rate, Avg Spindle speed, Length of cut in x, Length of cut in y
  - Simulated data: Depth of cut, Cutting strategy, (Volume of material cut), (Material cut in x), (Material cut in y), (Tool path strategy)
First, one must acquire the training data set.

### Table: Spindle Speed and Chip Load

<table>
<thead>
<tr>
<th>Level</th>
<th>Spindle Speed (RPM)</th>
<th>Chip Load (mm/tooth)</th>
<th>Depth of Cut (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1500</td>
<td>0.0254</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3000</td>
<td>0.0330</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>4500</td>
<td>0.0432</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>6000</td>
<td>0.0508</td>
<td></td>
</tr>
</tbody>
</table>

**Condensed and contextualized data**

<Input features – output response>
Description of training data set

Input:
- $x_1 \in \mathbb{R}$ Feed rate
- $x_2 \in \mathbb{R}$ Spindle speed
- $x_3 \in \mathbb{R}$ Depth of cut
- $x_4 \in \{1, 2, 3, 4\}$ Active tool axis ID (1 for $x$-axis, 2 for $y$-axis, 3 for $z$-axis and 4 for $x$-$y$ direction)
- $x_5 \in \{1, 2, 3\}$ Cutting strategy (1 for conventional, 2 for climbing and 3 for both)

$l \in \mathbb{R}$ Length of tool a tool path in NC code block

Output:
- $y = E/l \in \mathbb{R}$ Energy density (energy consumption per unit length of a tool path) in NC code block.

In total, 3,092 pairs of $x$ (machine operation feature vector) and $y$ (energy density) collected from the experiments. That is, $\{(x^i, y^i)|i = 1, ..., 3,092\}$ serve as the basis for this study.
Classifying training data set by machine operations

- **Cutting operations**: Face milling, Contouring, Slotting, Pocketing, Spiraling, Drilling
- **Non-cutting operations**: Air-cut in $x - y$ direction, Air-cut in $z$ direction, Rapid motion

<table>
<thead>
<tr>
<th>$n$ machine parameters</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_1^1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$x_2^1$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$x_n$</td>
<td>$x_n^1$</td>
</tr>
</tbody>
</table>

- **Face milling**
  \[ D_1 = \{(x^i, y^i) | i = 1, ..., m_1\} \]
  \[ \hat{y} = f_1(x) \]

- **Contouring**
  \[ D_2 = \{(x^i, y^i) | i = 1, ..., m_2\} \]
  \[ \hat{y} = f_2(x) \]

- **Cutting operations**
  \[ \cdots \]

- **Non-cutting operations**
  \[ D_q = \{(x^i, y^i) | i = 1, ..., m_q\} \]
  \[ \hat{y} = f_q(x) \]

- **Energy Prediction models for machine operations**

- **Training data**: Clustered input and output data for each machine operation
How to construct prediction function?

\[ D_q = \{(x^i, y^i)|i = 1, ..., m_q\} \quad \text{and} \quad \hat{y} = f_q(x) \]
Which learning algorithm to choose? : Gaussian Process (GP) regression

Gaussian Process
• model complex input and output relationships without the basis functions
• update model with new measurement data based on Bayesian framework
• estimate uncertainty in prediction
Given training data set for $q$th operation

$$D_q = \{(x^i, y^i); i = 1, ..., m_q\}$$

Assumption: measurements are corrupted with noise

$$y = f_q(x) + \epsilon, \epsilon \sim N(0, \sigma^2_{error})$$

Prior on the measured outputs

$$\begin{bmatrix} y^{1:m_q} \\ y \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & k \\ k^T & k(x, x) \end{bmatrix}\right)$$

Conditionalization on observed data

(Bayesian updating)

$$y | D_q \sim N\left(\mu(x|D_q), \sigma^2(x|D_q)\right)$$

$$\mu(x|D_q) = k^TK^{-1}y^{1:m_q}$$

$$\sigma^2(x|D_q) = k(x, x) - k^TK^{-1}k$$

$$k(x^i, x^j) = \tau^2\exp\left(-\frac{1}{2}(x^i - x^j)^T\text{diag}(\lambda)^{-2}(x^i - x^j)\right) + \sigma^2_{\epsilon}\delta_{ij}$$

$$k^T = (k(x^1, x), ..., k(x^{m_q}, x))$$

$$K_{ij} = k(x^i, x^j)$$
How GP constructs regression model from data?

True \( f(x) = x \sin(x) \)

Sampled without error

\[
n = 2 \quad \begin{bmatrix} y_1 \\ y_2 \\ y \\ y \\ y \\ k(x, x) \\ k(x, x) \\ k(x, x) \\ k(x, x) \\ k(x, x) \end{bmatrix} \sim N \left( 0, \begin{bmatrix} k & k^T \\ k^T & k(x, x) \end{bmatrix} \right)
\]
How GP constructs regression model from data?

True $f(x) = x \sin(x)$

Sampled without error

$n = 3$

$$\begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix} \sim N \left( 0, \begin{bmatrix} k & k \\ k^T & k(x, x) \end{bmatrix} \right)$$
Gaussian Process (simple example)

True $f(x) = x \sin(x)$

Sampled without error

$n = 4$

$$
\begin{bmatrix}
y^1 \\
y^2 \\
y^3 \\
y^4
\end{bmatrix} \sim N \left( 0, 
\begin{bmatrix}
\ddots & \ddots & k \\
\ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots \\
& \ddots & \ddots \\
k^T & k(x, x)
\end{bmatrix}
\right)
$$
How GP constructs regression model from data?

True \( f(x) = x \sin(x) \)

Sampled without error

\[
\begin{bmatrix}
  y^1 \\
  y^2 \\
  y^3 \\
  y^4 \\
  y^5 \\
\end{bmatrix} \sim N \left( \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  0 \\
\end{bmatrix}, \begin{bmatrix}
  k & k(x, x_1) & k(x, x_2) & \cdots & k(x, x_n) \\
  k^T & k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  k^T & k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \\
\end{bmatrix} \right)
\]
How GP constructs regression model from data?

True $f(x) = x \sin(x)$

Sampled without error

$n = 6$

$$
\begin{bmatrix}
y^1 \\
y^2 \\
y^3 \\
y^4 \\
y^5 \\
y^6 \\
y
\end{bmatrix} \sim N \left( 0, 
\begin{bmatrix}
\ddots & & & & & & \\
& \ddots & & & & & \\
& & \ddots & & & & \\
& & & \ddots & & & \\
& & & & \ddots & & \\
& & & & & \ddots & \\
0 & & & & & & k
\end{bmatrix}
\right)
$$
How GP constructs regression model from data?

True \( f(x) = x \sin(x) \)

Sampled with error:

\[
f(x) = x \sin(x) + \epsilon, \quad \epsilon \sim N(0, 1^2)
\]

\[
n = 2 \begin{bmatrix} y^1 \\ y^2 \\ y \end{bmatrix} \sim N \left( \begin{bmatrix} \square \square \\ \square \square \\ k \end{bmatrix}, \begin{bmatrix} k^T & k(x, x) \end{bmatrix} \right)
\]
How GP constructs regression model from data?

True $f(x) = x \sin(x)$

Sampled with error:

$$f(x) = x \sin(x) + \epsilon, \quad \epsilon \sim N(0, 1^2)$$

$$n = 3$$

$$\begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix} \sim N \begin{pmatrix} 0, & \begin{bmatrix} k_{y}^T & k \\ k & k(x, x) \end{bmatrix} \end{pmatrix}$$
How GP constructs regression model from data?

True \( f(x) = x\sin(x) \)

Sampled with error:

\[
f(x) = x\sin(x) + \epsilon, \quad \epsilon \sim N(0, 1^2)
\]

\[
\begin{bmatrix}
y^1 \\
y^2 \\
y^3 \\
y^4 \\
y
\end{bmatrix} \sim N \begin{pmatrix}
0, & k \\
k^T & k(x, x)
\end{pmatrix}
\]

\(n = 4\)
How GP constructs regression model from data?

True \( f(x) = x \sin(x) \)

Sampled with error:

\[
f(x) = x \sin(x) + \epsilon, \quad \epsilon \sim N(0, 1^2)
\]

\[
\begin{bmatrix}
y^1 \\
y^2 \\
y^3 \\
y^4 \\
y^5 \\
y
\end{bmatrix} \sim N \left( \begin{bmatrix} \\
\end{bmatrix}, 
\begin{bmatrix} 
& k \\
k^T & k(x, x)
\end{bmatrix} \right)
\]

\( n = 5 \)
How GP constructs regression model from data?

True $f(x) = x\sin(x)$

Sampled with error:

$$f(x) = x\sin(x) + \epsilon, \quad \epsilon \sim N(0, 1^2)$$

$n = 6$
Constructed energy density prediction function

Energy density prediction function $\hat{y} = f_1(x)$ for face milling

mean function $\mu(x|D_1)$ (5D function)

Standard deviation $\sigma(x|D_1)$ (5D function)

Prediction can be represented with bound:

$$[\mu(x|D_1) - \sigma(x|D_1), \mu(x|D_1) + \sigma(x|D_1)]$$

$x_2$
Spindle Speed (RPM)

$x_3$
Depth of cut (mm)
From density prediction to energy prediction

- Energy density prediction model for machine operation type $q$
  \[ \hat{y} = f_q(x) \]
  - Mean function $\mu(x|D_q)$
  - Standard deviation function $\sigma(x|D_q)$

- Energy consumption for $i$th NC code block performing machine operation type $q$:
  \[ \hat{E}^i = \mu_q(x^i|D_q) \times l^i \]
  \[ S^i = \sigma_q(x^i|D_q) \times l^i \]

- Energy consumption for NC code blocks performing machine operation type $q$:
  \( \hat{E}_q = \sum_{(x^i,y^i) \in D_q} \mu_q(x^i|D_q) \times l^i \)
  \[ S_q = \sqrt{\sum_{(x^i,y^i) \in D_q} \left( \sigma_q(x^i|D_q) \times l^i \right)^2} \]

- Energy consumption for the entire operations:
  \[ \hat{E} = \sum_{q=1}^{Q} \hat{E}_q \]
  \[ S = \sqrt{\sum_{q=1}^{Q} (S_q)^2} \]
  \( E \sim N(\hat{E}, S) \) : Probabilistic prediction
Can we predict how much energy the manufacturing machine will consume when machining the part? → **YES**

We test whether the prediction model prediction energy consumptions for machining parts

- with different geometry
- with different machine control parameters (in this case study, varying spindle speeds)

<table>
<thead>
<tr>
<th>Level of generalization</th>
<th>Used spindle speeds (RPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training parts 1~18</td>
<td>{1,500, 3,000, 4,500, 6,000}</td>
</tr>
<tr>
<td>Test part 1</td>
<td>{1,500, 3,000, 4,500}</td>
</tr>
<tr>
<td>Test part 2</td>
<td>{1,700, 2,800, 4,300}</td>
</tr>
<tr>
<td>Test part 3</td>
<td>{2,125, 2,400, 3,750}</td>
</tr>
</tbody>
</table>

Energy consumption $E \sim N(\hat{E}, S)$
Prediction results for test parts

Energy density prediction for face milling $\hat{y} = f_1(x)$
Prediction results for test parts

Energy consumptions for each NC code block

Error rates

<table>
<thead>
<tr>
<th>No. of data</th>
<th>Averaged block duration (sec)</th>
<th>RAЕ (%)</th>
<th>Predicted total energy (KJ)</th>
<th>Measured total energy (KJ)</th>
<th>Standard deviation (KJ)</th>
<th>RTE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>188</td>
<td>10.27</td>
<td>13.004</td>
<td>22.492</td>
<td>0.434</td>
<td>3.702</td>
</tr>
<tr>
<td>Test 2</td>
<td>188</td>
<td>9.82</td>
<td>15.210</td>
<td>21.928</td>
<td>0.441</td>
<td>0.290</td>
</tr>
<tr>
<td>Test 3</td>
<td>188</td>
<td>9.70</td>
<td>23.143</td>
<td>21.747</td>
<td>0.477</td>
<td>3.192</td>
</tr>
</tbody>
</table>

- \( \text{RAE} = \frac{\sum\{i \in \text{NC blocks}\} |\hat{E}^i - E^i|}{\sum\{i \in \text{NC blocks}\} E^i} \)
- \( \text{RTE} = \frac{|E - \hat{E}|}{\hat{E}} \)
Future work and Conclusion

- **Past Experience**
- **Target Machine Tool**
- **External sensors**

Process Parameters:
- **Optimization Process Parameters**

**Knowledge Archive**

**MTConnect Agent**

**Data Processor**

**Data Buffer**

**Knowledge Extraction Agent**

**Adaptive Machine Learning**

**Real time data acquisition and processor**

- **Real time data acquisition**
- **Continuous machine monitoring and control without saving big data**
- **Adaptive machine learning**
Acknowledgements and disclaimer

• The authors acknowledge the support in part of the Smart Manufacturing Systems Design and Analysis Program at the National Institute of Standards and Technology (NIST), Grant Numbers 70NANB12H225 and 70NANB12H273 awarded to University of California, Berkeley, and to Stanford University respectively. In addition, the authors appreciate the support of the Machine Tool Technologies Research Foundation (MTTRF) and System Insights for the equipment used in this research.

• Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology (NIST); nor does it imply that the products identified are necessarily the best available for the purpose. Further, any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIST or any other supporting U.S. government or corporate organizations.
Supplementary slides
Limitations in GP

Issues:
• The amount of monitoring data increases with time.
• The extractable knowledge is not necessarily proportional to the amount of data.
• Constructing GP model is computationally expensive $O(m^3)$, $m$ is number of data points.

Requirements:
• The amount of data storage should be minimized while maximizing the knowledge extraction
  → The extracted knowledge should be updated with new measurement data to account for the time varying characteristics of a target machine, i.e., tool wear or aging.
Collective Gaussian Process

• Different types of knowledge are processed and retained by different region of brain.

• The knowledge is updated with new information, i.e., outdated knowledge is replaced with new one.

Online collective GP is composed of

• $p_i(x)$: input feature distribution constructed by Gaussian Mixture Model using local data set. (domain of knowledge).

• $h_i(x)$: regression model mapping input to output constructed by GP using local data set. (content of knowledge).
Given $N$ pairs of energy prediction function and input features’ PDF, $\{(h_i(x), p_i(x))\} | i = 1, 2, \ldots, N$ where the weighting coefficient $w_i$ for the $i$th local energy prediction function is determined as

$$w_i = \frac{p_i(x^{new})}{\sum_{j}^{N} p_j(x^{new})}$$
Gaussian Mixture Model (GMM)

Modeling a probability density function as a combination of \( K \) Gaussian components

\[
p(x; \varphi, \mu, \Sigma) = \sum_{k=1}^{K} g_k(x; \mu_k, \Sigma_k) \varphi_k
\]

\( K \): number of GPDFs
\( \varphi = \{\varphi_1, ..., \varphi_K\} \): set of weights
\( \mu = \{\mu_1, ..., \mu_K\} \): set of mean vectors
\( \Sigma = \{\Sigma_1, ..., \Sigma_K\} \): set of covariance matrices

For an feature vector \( x = (x_1, x_2, ..., x_n) \), the \( k \)th component density is of a form of Gaussian

\[
g_k(x; \mu_k, \Sigma_k) = N(x; \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_k|}} \exp \left(-\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right)
\]

\( \varphi_k \): (mixture) weight for \( k \)th Gaussian component (\( \sum_{k=1}^{K} \varphi_k = 1 \))
\( \mu_k \): mean vector for the \( k \)th Gaussian component
\( \Sigma_k \): covariance matrix for the \( k \)th PDF
Gaussian Mixture Model (GMM)

Simple example

\[
p(x) = \sum_{k=1}^{K} g_k(x) \varphi_k
\]

Weighted sum of Gaussian PDFs

\[g_1(x)\quad g_2(x)\quad g_3(x)\]
Numerical results

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Computational time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RAE (%)</td>
<td>RTE (%)</td>
</tr>
<tr>
<td>Collective GP</td>
<td>15.45</td>
<td>-0.41</td>
</tr>
<tr>
<td>Global GP</td>
<td>17.81</td>
<td>0.72</td>
</tr>
</tbody>
</table>

The comparison of prediction accuracies between the collective GP regression and other local GP regression methods

<table>
<thead>
<tr>
<th>Weighting methods</th>
<th>RAE (%)</th>
<th>RTE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability by GMM</td>
<td>17.81</td>
<td>0.72</td>
</tr>
<tr>
<td>Variance by GP</td>
<td>18.40</td>
<td>0.14</td>
</tr>
<tr>
<td>Geometric distance to center</td>
<td>42.41</td>
<td>24.02</td>
</tr>
<tr>
<td>Average</td>
<td>48.47</td>
<td>39.50</td>
</tr>
</tbody>
</table>
Prediction with blind test parts

Part design → NC code → Target machine → Machined part

Simulator

Machine parameters

$$\begin{bmatrix}
  x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\
  x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)}
\end{bmatrix}$$

$$\{f_q(x) | q = 1, \ldots, Q\} \xrightarrow{} E \sim N(\hat{E}, S)$$

Energy consumption
Energy prediction on test data set randomly selected from training data set.
Data acquisition & simulator for this study

**Hardware**

- **Fanuc Controller**:
  Collect machine control parameters

- **System Insights High Speed Power Meter (HSPM)**:
  Collect power time series

**Data types**

- **Direct data**:
  Directly measured from hardware

- **Derived data**:
  Derived through simple calculations, e.g., mean and tool path computation.

- **Simulated data**:
  Obtained by simulation, accounting for workpiece geometry