Improving an Ecosystem Model Using Earth Science Data

Kazumi Saito (NTT Communication Science Laboratories)
Pat Langley (CSLI Stanford University)
Trond E. Grenager (Stanford University)
Introduction

• Developing computational methods for discovering knowledge in communicable forms.

• Improving CASA using observed data.

• CASA: an existing computational model of aspect of the Earth ecosystem developed by Christopher Potter and his colleagues at NASA Ames.
Some Equations

$NPPc$: net primary production.

$$NPPc = \max (0, E \times IPAR)$$

$E$: value of maximum possible photosynthetic efficiency under temperature and moisture stress scalars.

$$E = e_{max} \times T1 \times T2 \times W$$

$IPAR$: converter for intercepted photosynthetically active radiation by the vegetation cover.

$$IPAR = FPAR_{FAS} \times Solar \times sol\_conv \times 0.5$$
General Problem

• Revisions to the model must be consistent with existing knowledge of Earth science and, ideally, retain similarity to the current model.

• Our research involves attempting to improve the CASA model’s predictive accuracy.
Outline of Approach

• Transforming the equations into a neural network
• Revising weights in that network
• Transforming the network back into equations

\[ NPPc = f(\cdots) \]

\[ \cdots \]

\[ \cdots \]

original equations

revised equations

\[ NPPc = g(\cdots) \]
Some Types of Neural Networks

Standard (sigma-sigma) net:

$$ \sum w_j f_j \left( \sum w_{jk} x_k \right) $$

Sigma-pi net (generalized polynomial):

$$ \sum w_j \prod x_k^{w_{jk}} = \sum w_j \exp \left( \sum w_{jk} \ln x_k \right) $$

Pi-sigma net (this talk):

$$ \prod w_j f_j \left( \sum w_{jk} x_k \right) $$
Transforming Equations

Stress scalars

NPPc

E

IPAR

PET

eet

Topt

SR_FAS

PET_TW_M

tempc

FAS_NDVI

srmin

srmax

Intrinsic property

sol_conv

umd_veg

Variable

Observed Variable

Known Constant
Stress Scalars

Original equations:

\[ E = e_{\text{max}} \times T1 \times T2 \times W \]

\[ T1 = 0.8 + 0.02 \times T_{\text{opt}} - 0.0005 \times T_{\text{opt}}^2 \]

\[ = 1 - (-0.4472 + 0.0224 \times T_{\text{opt}})^2 \]

\[ T2 = 1.1814 \times \frac{1}{1 + \exp(0.2 \times (-10 + T_{\text{opt}} - \text{tempc}))} \]

\[ \times \frac{1}{1 + \exp(0.3 \times (-10 - T_{\text{opt}} + \text{tempc}))} \]

\[ W = 0.5 + 0.5 \times \left( \frac{e_{\text{et}}}{\text{PET}} \right) \]
Transformation into Network

\[ E = e \max \times T1 \times T2 \times W = w_0 \times \prod f_i \]

\[ T1 = f_1(x) = 1 - x^2 = f_1(w_{11} + w_{12} \times Topt) \]
\[ = 1 - (-0.4472 + 0.0224 \times Topt)^2 \]

\[ T2 = f_2(x) = f_{21}(x) \times f_{22}(x) = \frac{1}{1 + \exp(-x)} \times \frac{1}{1 + \exp(-x)} \]
\[ f_{21}(x) = f_{21}(w_{21} + w_{22} (Topt - tempc)) \]
\[ = \frac{1}{1 + \exp(2 - 0.2 \times (Topt - tempc))} \]

\[ W = f_3(x) = x = f_3\left( w_{31} + w_{32} \frac{eet}{PET} \right) = 0.5 + 0.5 \times \frac{eet}{PET} \]
Intrinsic Values for Vegetation Type

$FPAR\_FAS$: fraction of absorbed photosynthetically active radiation by the vegetation cover

$$FPAR\_FAS = \min\left(\frac{SR\_FAS - srmin}{SRDIFF}, 0.95\right)$$

$$\approx \frac{1}{SRDIFF} \times (SR\_FAS - srmin)$$

$SRDIFF$: map from the ground cover to an $srmax-srmin$ value
Transformation into Network

\[
\frac{SR_{FAS} - srmin}{SRDIFF} = \exp\left(\log(SR_{FAS} - srmin) - \log(SRDIFF)\right)
\]

\[-\log(SRDIFF) = \sum_{i=1}^{13} v_i \times umd\_veg\_i\]

\(v_i\) : weight in neural network

\(umd\_veg\_i = \begin{cases} 1 & \text{if } umd\_veg = i \\ 0 & \text{otherwise} \end{cases}\)
### Revising weights in Networks

<table>
<thead>
<tr>
<th>Supervised Learning</th>
<th>Step-length</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed (constant)</td>
<td>variable</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Direction</th>
<th>1st-order</th>
<th>2nd-order</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP, etc.</td>
<td>Newton method</td>
<td>Silva-Almeida algorithm, etc.</td>
</tr>
<tr>
<td>SCG, OSS, BPQ, etc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd-order learning algorithm</th>
<th>Applicability to large-scale problems</th>
<th>Performance with inaccurate step-length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss-Newton method</td>
<td>✗</td>
<td>○</td>
</tr>
<tr>
<td>quasi-Newton method</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>conjugate gradient method</td>
<td>○</td>
<td>✗</td>
</tr>
</tbody>
</table>
BPQ Algorithm

• The search direction is calculated on the basis of partial BFGS update.

• The step-length is calculated by using a second-order approximation.
Demonstration Problem

Sample set

<table>
<thead>
<tr>
<th>Sample</th>
<th>y</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.73</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.95</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>0.99</td>
<td>5</td>
</tr>
</tbody>
</table>

Input unit

Hidden unit

Output unit
Learning Neural Network: Result

Squared error

- BPQ
  (2nd-order method)

- BP + momentum
  (1st-order method)
The RMSE of the original model was reduced by 15 percent, as measured using cross validation.

\[
RMSE = \sqrt{\frac{\sum_{samples} (NPP_{observed} - NPP_{predicted})^2}{\text{number of samples}}}
\]
The intrinsic values associated with vegetation types obtained in this way were consistently lower.
Transforming Network

Step 1. Quantize \( \{ \exp \left( \sum_{kl} v_{kl} q_{kl}^{(n)} \right) : n = 1, \ldots, N \} \) by using a clustering method.

Step 2. Determine an adequate number of rules by using cross-validation.

Clustering Analysis

Number of clusters

- apparent RMSE
- LOO CV RMSE
Evaluating Experimental Result

![Bar Chart](image)

- **Initial RMSE**
- **Appar**
- **LOO CV RMSE**

**Before Clustering**

**After Clustering**
Obtained Decision Tree

\[ t_8 = 1 : 0 \ (10.0) \]
\[ t_8 = 0 : \]
\[ \quad t_9 = 1 : 1 \ (58.0) \]
\[ \quad t_9 = 0 : \]
\[ \quad \quad t_7 = 1 : 1 \ (46.0) \]
\[ \quad \quad t_7 = 0 : \]
\[ \quad \quad \quad t_{11} = 1 : 1 \ (11.0) \]
\[ \quad \quad \quad t_{11} = 0 : \]
\[ \quad \quad \quad \quad t_1 = 0 : 2 \ (168.0/1.0) \]
\[ \quad \quad \quad \quad t_1 = 1 : 1 \ (10.0) \]
Clustered Intrinsic Values

![Clustered Intrinsic Values Chart]

- **Initial**
- **Obtained**
Conclusion

• This talk described an approach to improving the predictive accuracy of the existing ecosystem model.

• In the experiments, we can reduce the mean squared error of the original model by 15 percent, as measured using cross validation.

• In the future, we’ll carry out further experiments along this direction.