Interpreting and Explaining Deep Neural Networks: A Perspective on Time Series Data – Part 3/3

Jaesik Choi

Explainable Artificial Intelligence Center
Graduate School of Artificial Intelligence
KAIST

Some slides courtesy of Eamonn Keogh
Interpreting and Explaining Deep Neural Networks: A Perspective on Time Series Data

Agenda (150 min)

Overview to Explainable Artificial Intelligence (XAI) – 15 min
Input Attributions Methods for Deep Neural Networks – 35 min
Interpreting Inside of Deep Neural Networks – 50 min

Explainable Models for Time Series Data – 50 min
  - Important Questions on Mining and Learning Time Series Data
  - Visualizing Deep Temporal Neural Networks
  - Clustered Pattern of Highly Activated Period (CPHAP)
  - Automatic Statistician/Relational Automatic Statistician (Bayesian Approaches)
1. Have we ever seen a pattern that looks just like this?
2. Are there any repeated patterns in my data?
3. What are the three most unusual days in this three month long dataset?
4. Is there any pattern that is common to these two time series?
5. How do these two time series differ in terms of alignment?
6. Find the most conserved pattern that happens at least once every two days in this two week long dataset.
7. If you had to summarize this long time series with just two shorter examples, what would they be?
8. Are there any patterns that appear as time reversed versions of themselves in my data?
9. When does the regime change in this time series?
10. How can I compare these time series of different lengths?
11. Are there any patterns that repeat in my data, but at two distinct lengths?
12. Have we ever seen a multidimensional pattern that looks just like this?
13. How do I quickly search this long dataset for this pattern, if an approximate search is acceptable?
14. How can I optimize similarity search in a long time series?
15. What is most likely to happen next?
16. What is the right length for motifs in this dataset?
17. I need to find motifs faster! Part I
18. I need to find motifs faster! Part II
19. Have we ever seen a pattern that looks just like this, but possibly at a different length?
20. How can I know which of these two classification approaches is best for time series?
21. Are there any evolving patterns in this dataset (time series chains)
22. (pending)
Find the subsequences having very high similarity to each other.
The electrical penetration graph or EPG is a system used by biologists to study the interaction of insects with plants.

15 minutes of EPG recorded on Beet Leafhopper

As a bead of sticky secretion, which is by-product of sap feeding, is ejected, it temporarily forms a highly conductive bridge between the insect and the plant.

Slides courtesy of Eamonn Keogh
More motifs reveal different feeding patterns of Beet Leafhopper.

Slides courtesy of Eamonn Keogh
The dataset is an hour of EOG (eye movement) data of a sleeping patient, sampled at 100 Hz.

Note that there may be more examples of each motif!

Slides courtesy of Eamonn Keogh

Are there any repeated patterns in my data? – Motif Search
Here, we are interested in finding Temporal Motifs Trained in Deep Temporal Neural Networks
Temporal Neural Networks: MLP vs FCN vs ResNet

### Experimental Results on UCR dataset

Experimental Results


<table>
<thead>
<tr>
<th></th>
<th>Err Rate</th>
<th>DTW</th>
<th>COTE</th>
<th>MCNN</th>
<th>BOSSVS</th>
<th>PROP</th>
<th>BOSS</th>
<th>SE1</th>
<th>TSBF</th>
<th>MLP</th>
<th>FCN</th>
<th>ResNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Win</td>
<td></td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>MPCE</td>
<td></td>
<td>0.0397</td>
<td>0.0226</td>
<td>0.0241</td>
<td>0.0330</td>
<td>0.0304</td>
<td>0.0256</td>
<td>0.0302</td>
<td>0.0335</td>
<td>0.0407</td>
<td><strong>0.0219</strong></td>
<td>0.0231</td>
</tr>
</tbody>
</table>

https://github.com/cauchyturing/UCR_Time_Series_Classification_Deep_Learning_Baseline
Residual Network [ResNet, He et. al., 2015]

He et. al., 2015

Residual learning

\[ x_\ell = H_\ell(x_{\ell-1}) + x_{\ell-1} \]

Comparison of Resnet

3.6% of error in ImageNet Challenge, 2015
Recurrent Convolutional Neural Layers [RCNN, Liang and Hu, 2015]

Liang and Hu, 2015

\[ x_l = x_{l-1} + H_l(x_{l-1}) + H_l(H_l(x_{l-1}))) + H_l(H_l(H_l(x_{l-1}))) \]
Hand Start

First Digit Touch

Lift off

Replace

Both Released

* Joint work with Azamatbek Akhmedov

RCNN on EEG Analysis

Luciw et. al., 2014
One chunk: Data: 3584,32

* Joint work with Azamatbek Akhmedov

RCNN on EEG Analysis
Applying RCL

<table>
<thead>
<tr>
<th>Layer type</th>
<th>Size</th>
<th>Output shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolutional</td>
<td>256 1×9 filters</td>
<td>(64, 256, 1, 3584)</td>
</tr>
<tr>
<td>Max pooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCL: (1,896)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max pooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCL: (1,224)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max pooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCL: (1,56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max pooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCL: (1,14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max pooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully connected</td>
<td>1792×6</td>
<td>(64, 6)</td>
</tr>
</tbody>
</table>

97.687%

RCNN on EEG Analysis
256 1x9 filters
Example: Hand Start

RCNN on EEG Analysis
Example: Hand Start

RCNN on EEG Analysis
Example: Hand Start

RCNN on EEG Analysis
Example: First Digit Touch

RCNN on EEG Analysis
Example: First Digit Touch

RCNN on EEG Analysis
Example: First Digit Touch

RCNN on EEG Analysis
Example: First Digit Touch

RCNN on EEG Analysis
Example: Replace

RCNN on EEG Analysis
Example: Replace

RCNN on EEG Analysis
Example: Replace

RCNN on EEG Analysis
Example: Replace
How can we separate time series data into semi-global representative parts without hand-crafted segmentation labels for interpreting?

Clustered Pattern of Highly Activated Period: Motivation

S. Cho et al., 2020
Clusters from Layer 1

Clusters from Layer 2

Clusters from Layer 3

Input

Classification

CPHAP

Clustered Pattern of Highly Activated Period

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period (HAP)

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period: Results

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period: Results

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period: Results

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period: Results

S. Cho et al., 2020
Clustered Pattern of Highly Activated Period: Results

S. Cho et al., 2020
Experiment 1: Different Network Structure (ResNet)

S. Cho et al., 2020
Experiment 2: Different Filter Size

S. Cho et al., 2020
Experiment 3: Sequences of test data with CPHAP of train data

S. Cho et al., 2020
Experiment 4: Visual Comparison among XAI methods

S. Cho et al., 2020
Experiment 5: Perturbating with unimportant area

S. Cho et al., 2020

https://clusteredpattern.github.io/pages/
Automation of Knowledge Work

**Work activity summary:** Finance and insurance

*Grey lines represent average per activity across all sectors.*

<table>
<thead>
<tr>
<th>Activity type</th>
<th>Time spent by activity</th>
<th>Automation potential, percentage of time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictable physical work</td>
<td>22%</td>
<td>54%</td>
</tr>
<tr>
<td>Data processing</td>
<td>14%</td>
<td>67%</td>
</tr>
<tr>
<td>Data collection</td>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>Unpredictable physical work</td>
<td>9%</td>
<td>14%</td>
</tr>
<tr>
<td>Stakeholder interactions</td>
<td>14%</td>
<td>12%</td>
</tr>
<tr>
<td>Applying expertise</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>Managing others</td>
<td>6%</td>
<td>7%</td>
</tr>
</tbody>
</table>

**SOURCE:**
https://public.tableau.com/profile/mckinsey.analytics#!/vizhome/AutomationBySector/WhereMachinesCanReplaceHumans
SAN JOSE, Calif. (AP) — Adobe Systems Inc. (ADBE) on Tuesday reported fiscal
third-quarter profit of $419.6 million.
The San Jose, California-based company said it had profit of 84 cents per share.
Earnings, adjusted for one-time gains and costs, were $1.10 per share.

Adobe shares have climbed 52 percent since the beginning of the year. In the
final minutes of trading on Tuesday, shares hit $156.61, an increase of 57
percent in the last 12 months.

This story was generated by Automated Insights
(http://automatedinsights.com/ap) using data from Zacks Investment Research
Sonoma County Little Leagues (Falcons vs Mustangs)

Anthony T got it done on the bump on the way to a win. He allowed two runs over 2-1/3 innings. He struck out four, walked two, and surrendered no hits. Anders Mathison ended up on wrong side of the pitching decision, charged with the loss. He lasted just two innings, walked two, struck out one, and allowed four runs.

Automated generated by Quill, Narrative Science
Each of 45 respondents read a game recap article and decide whether or not the text had been written by a journalist or by a computer.
Automated Insights is acquired by Vista for $80 million (Feb. 2015).

Narrative Science get funded $43.4 million, so far.

...
SAN JOSE, Calif. (AP) -- Adobe Systems Inc. (ADBE) on Tuesday reported fiscal third-quarter profit of $419.6 million. The San Jose, California-based company said it had profit of 84 cents per share. Earnings, adjusted for one-time gains and costs, were $1.10 per share.

Adobe shares have climbed 52 percent since the beginning of the year. In the final minutes of trading on Tuesday, shares hit $156.61, an increase of 57 percent in the last 12 months.

This story was generated by Automated Insights (http://automatedinsights.com/ap) using data from Zacks Investment Research
Sonoma County Little Leagues (Falcons vs Mustangs)

Anthony T got it done on the bump on the way to a win. He allowed two runs over 2-1/3 innings. He struck out four, walked two, and surrendered no hits.

Anders Mathison ended up on wrong side of the pitching decision, charged with the loss. He lasted just two innings, walked two, struck out one, and allowed four runs.

Generated by Quill, Narrative Science
The deeper challenge lies not in generating copy, but in finding the most pertinent meaning in a given dataset.

“It’s not just about converting numbers to language.”

“Those numbers need context”
Finding Context in Time Series Data
Descriptive prediction of time series

Problem
Descriptive prediction of time series

Problem
Descriptive prediction of time series

Problem
Function $G\mathcal{P}(\mu(x), k(x, x'))$:

- **Mean function**
  \[ \mu(x) = \mathbb{E}(f(x)) \]

- **Covariance kernel function**
  \[ k(x, x') = \text{Cov}(f(x), f(x')) \]

$f(x) \sim G\mathcal{P}(\mu(x), k(x, x'))$
\[ f(x) \sim \mathcal{GP}(\mu(x), k(x, x')) \]

**Mean function**
\[ \mu(x) = \mathbb{E}(f(x)) \]

**Covariance kernel function**
\[ k(x, x') = \text{Cov}(f(x), f(x')) \]

**Function evaluations**
\[ [f(x_1), \ldots, f(x_N)] \sim \mathcal{N}(\mu, \Sigma) \]

**Gaussian Processes (GP)**

\[ \mu = [\mu(x_1), \ldots, \mu(x_N)] \]

**Mean vector**
\[ \Sigma_{ij} = k(x_i, x_j) \]

**Covariance matrix**

**Multivariate Gaussian**
\[
f(x) \sim GP(\mu(x), k(x, x'))
\]
<table>
<thead>
<tr>
<th>Base kernel</th>
<th>Encoding function</th>
<th>Kernel function</th>
<th>Parameters</th>
<th>Example kernel function shape</th>
<th>Example encoded functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIN($x, x'$)</td>
<td>Linear function</td>
<td>$\sigma^2(x - \ell)(x' - \ell)$</td>
<td>$\sigma, \ell$</td>
<td><img src="example1.png" alt="kernel" /></td>
<td><img src="example2.png" alt="encoded functions" /></td>
</tr>
<tr>
<td>SE($x, x'$)</td>
<td>Smooth function</td>
<td>$\sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right)$</td>
<td>$\sigma, \ell$</td>
<td><img src="example3.png" alt="kernel" /></td>
<td><img src="example4.png" alt="encoded functions" /></td>
</tr>
<tr>
<td>PER($x, x'$)</td>
<td>Periodic function</td>
<td>In appendix</td>
<td>$\sigma, \ell, p$</td>
<td><img src="example5.png" alt="kernel" /></td>
<td><img src="example6.png" alt="encoded functions" /></td>
</tr>
</tbody>
</table>
(1) Encode characteristic

\[ f(x) \sim \mathcal{GP}(\mu(x), k(x, x')) \]

Find appropriate kernel

Multi-kernel Learning
\[f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))\]

(1) Encode characteristic

Find appropriate kernel

(2) Compose new kernel (appendix)

If \(g(x) \sim \mathcal{GP}(0, k_g), h(x) \sim \mathcal{GP}(0, k_h)\) and \(g(x) \perp h(x)\)

, then

\[g(x) + h(x) \sim \mathcal{GP}(0, k_g + k_h)\]

\[g(x) \times h(x) \sim \mathcal{GP}(0, k_g \times k_h)\]

The Automatic Statistician

*Automatic Bayesian Covariance Discovery (http://www.automaticstatistician.com/)

Ghahramani, 2015
<table>
<thead>
<tr>
<th>Op.</th>
<th>Concept</th>
<th>Params</th>
<th>Example</th>
<th>Example kernel function shape</th>
<th>Example encoded functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Addition Superposition OR operator</td>
<td>N/A</td>
<td>SE + PER</td>
<td><img src="image" alt="Example kernel function shape" /></td>
<td><img src="image" alt="Example encoded functions" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LIN + PER</td>
<td><img src="image" alt="Example kernel function shape" /></td>
<td><img src="image" alt="Example encoded functions" /></td>
</tr>
<tr>
<td>×</td>
<td>Multiplication AND operator</td>
<td>N/A</td>
<td>SE × PER</td>
<td><img src="image" alt="Example kernel function shape" /></td>
<td><img src="image" alt="Example encoded functions" /></td>
</tr>
</tbody>
</table>

The Automatic Statistician: Kernel Composition

Grosse et. al., 2012
Kernel Composition: Generate Data from Models

Kernel Composition & Covariance Decomposition
Kernel Composition: Generate Data from Models

\[ + \approx \]

Covariance Decomposition: Learn Explainable Models from Data

\[ \approx + \]
Iteratively select best model (structure $k$, parameter $\theta$)

1. **Expand**: the current kernel
2. **Optimize**: conjugate gradient descent
3. **Select**: the best kernel in the level (greedy)
4. **Iterate**: get back to (1) for the next level

---

**The Automatic Statistician: Greedy Kernel Search**

Duvenaud et. al., 2014
The Automatic Statistician: A Sample Report

Lloyd et. al., 2014
The Automatic Statistician: A Sample Report

Lloyd et. al., 2014
13 regression datasets

The Automatic Statistician: Extrapolation Performance

Lloyd et. al., 2014
Challenge: The Automatic Statistician
Incorporating Global Changes

Adjusted Close of General Electronics

Linear function
decrease x/week

Smooth function
Length scale: y weeks

Rapidly varying
smooth function
Length scale: z hours

9/11, 2001
Challenge: The Automatic Statistician
Incorporating Global Changes

Adjusted Close of General Electronics

Linear function
decrease $x$/week

Smooth function
Length scale: $y$ weeks

Rapidly varying smooth function
Length scale: $z$ hours

9/11, 2001

No Description on Influence of 9/11
Challenge: The Automatic Statistician

Q: How about handling multiple time series?

- Exploit multiple time series
- Find global descriptions
- Hope better predictive performance
Descriptive prediction of *multiple* time series

- **Constant function**
  - Sudden drop btw 9/12/01 ~ 9/15/01

- **Smooth function**
  - Length scale: $y$ weeks

- **Rapidly varying smooth function**
  - Length scale: $z$ hours

**Problem (Our research)**
A Generalized Multi Kernel Learning

\[ P(D | \mathcal{M}) = P(D | \mathcal{GP}(0, k(x, x'; \theta))) \]
Model: Semi-Relational Kernel Learning

\[ P(D|\mathcal{M}) = \prod_{j=1}^{M} P(d_j|\mathcal{GP}(0, \sigma_j \times k(x, x'; \theta) + k_j(x, x'; \theta_j))) \]

Hwang et al., 2016
<table>
<thead>
<tr>
<th>Descriptions</th>
<th>Graphs (normalized)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 adjusted close of stock figures (2001 ~ 2002)</td>
<td><img src="image1.png" alt="Graph" /></td>
<td>GE, MSFT, XOM, PFE, C, WMT, INTC, BP, AIG</td>
</tr>
<tr>
<td>6 US housing price indices (2003 ~ 2013)</td>
<td><img src="image2.png" alt="Graph" /></td>
<td>New York, Los Angeles Chicago, Phoenix, San Diego, San Francisco</td>
</tr>
<tr>
<td>4 emerging market currency exchanges (2016)</td>
<td><img src="image3.png" alt="Graph" /></td>
<td>Indonesian - IDR, Malaysian - MYR, South African - ZAR, Russian - RUB</td>
</tr>
</tbody>
</table>

**Experiments on Financial Data Sets**

Hwang et al., 2016
US stock market values suddenly drop after US 9/11 attacks.

Currency exchange is affected by FED’s policy change in interest rates around middle Sep 2015.

Qualitative Results

Hwang et al., 2016
Quantitative Results

Hwang et al., 2016
<table>
<thead>
<tr>
<th>Data set</th>
<th>Negative log likelihood</th>
<th>Bayesian Information Criteria</th>
<th>Root mean square error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CKL</td>
<td>RKL</td>
<td>SRKL</td>
</tr>
<tr>
<td>STOCK3</td>
<td>332.75</td>
<td>311.84</td>
<td>304.05</td>
</tr>
<tr>
<td>STOCK6</td>
<td><strong>972.00</strong></td>
<td>1007.09</td>
<td>988.14</td>
</tr>
<tr>
<td>STOCK9</td>
<td>1776.31</td>
<td>1763.96</td>
<td><strong>1757.11</strong></td>
</tr>
<tr>
<td>HOUSE2</td>
<td>264.69</td>
<td>304.29</td>
<td>310.38</td>
</tr>
<tr>
<td>HOUSE4</td>
<td>594.79</td>
<td><strong>586.81</strong></td>
<td>1249.82</td>
</tr>
<tr>
<td>HOUSE6</td>
<td><strong>849.64</strong></td>
<td>891.09</td>
<td>1495.40</td>
</tr>
<tr>
<td>CURRENCY4</td>
<td><strong>578.35</strong></td>
<td>617.77</td>
<td>693.76</td>
</tr>
</tbody>
</table>

STOCK3 = \{GE, MSFT, XOM\}  
STOCK6 = STOCK3 + \{PFE, C, WMT\}  
STOCK9 = STOCK6 + \{INTC, BP, AIG\}  
HOUSE2 = \{NY, LA\}  
HOUSE4 = HOUSE2 + \{Chicago, Phoenix\}  
HOUSE6 = HOUSE4 + \{San Diego, San Francisco\}  
CURRENCY4 = \{IDR, MYR, ZAR, RUB\}  

Quantitative Results  
Hwang et al., 2016
An Automatically Generated Report
An Automatically Generated Report
An Automatically Generated Report
An automatically generated report for the dataset: GE
Relational version

2.6 Component 6: A constant. This function applies from 12 Sep 2001 until 15 Sep 2001

This component is constant. This component applies from 12 Sep 2001 until 15 Sep 2001.
This component explains 100.0% of the residual variance; this increases the total variance explained from 95.2% to 100.0%. The addition of this component increases the cross validated MAE by 0.67% from 0.87 to 0.87. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

Figure 1: Raw data (left) and model posterior with extrapolation (right)
Challenges: Selective Kernel Search
Q: Can we selectively search over time series?
Indian Buffet Processes (IBP) + Gaussian Processes
(Nonparametric Clustering)  (Nonparametric Regression)

Discovering Explainable Latent Covariance Structures for Multiple Time Series
Tong, Choi, 2018
South African Rand and Indonesian Rupiah and Malaysian Ringgit and Russian Rouble share the following properties

→ This component is a smooth function with a typical lengthscale of 6.4 days. This component applies until Sep. 15th 2015 and from Sep. 17th 2015 onwards.

Indonesian Rupiah and Malaysian Ringgit and Russian Rouble share the following properties

→ This component is linearly increasing.
Discovering Explainable Latent Covariance Structures for Multiple Time Series – Version II

Tong et al., 2018
Discovering Explainable Latent Covariance Structures for Multiple Time Series

Tong et al., 2018
I read annual reports of the company I'm looking at and I read the annual reports of the competitors - that is the main source of material.
I read annual reports of the company I'm looking at and I read the annual reports of the competitors - that is the main source of material.

— Warren Buffett —
Future: Finding Explanation from Reports
Read the Report and Explain It

Multiple Stocks

Bayesian Learning

Annual Reports

Deep Learning

Report

Prediction
- Automated data collection and processing soon will change our daily life.

- Automated narrative generation methods/frameworks may have widespread applications such as finance and media.

- Compositions of explainable models would generate more human understandable descriptions of data.

- Reading and Explaining Articles (e.g., Annual Report) would greatly help to improve the prediction accuracy in the future.