TACTiS
Transformer-Attentional Copula for Time Series

Alexandre Drouin
Étienne Marcotte
Nicolas Chapados
Objective

**Goal:** Infer the joint distribution of masked time points, given the observed time points

Very general: forecasting, interpolation, or arbitrary patterns
Objective

**Goal:** Infer the joint distribution of masked time points, given the observed time points

Very general: **forecasting**, interpolation, or arbitrary patterns
Objective

**Goal:** Infer the joint distribution of masked time points, given the observed time points

Very general: forecasting, **interpolation**, or arbitrary patterns
Objective

Goal: Infer the joint distribution of masked time points, given the observed time points

Very general: forecasting, interpolation, or arbitrary patterns
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.

**Encoder:** each point in each time series is a token.
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.

Encoder: each point in each time series is a token
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.

Decoder: a copula-based autoregressive decoder

Theorem (Sklar): any joint distribution can be expressed as a combination of two components:

1. **Marginal** distribution of each variable
2. **Copula**: joint distribution on the unit cube

Why?
• Interpretability
• Robustness to distribution shifts
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.

Decoder: a copula-based autoregressive decoder

Theorem (Sklar): any joint distribution can be expressed as a combination of two components:

1. Marginal distribution of each variable
2. Copula: joint distribution on the unit cube

Why?
- Interpretability
- Robustness to distribution shifts
Overview of the model

TACTiS is an encoder-decoder model, similar to standard transformers.

Decoder: a copula-based autoregressive decoder

Theorem (Sklar): any joint distribution can be expressed as a combination of two components:

1. Marginal distribution of each variable
2. Copula: joint distribution on the unit cube

Why?
• Interpretability
• Robustness to distribution shifts
Decoding using attentional copulas
Decoding using attentional copulas

Memory: observed and previously decoded tokens

Sample

0 1
Decoding using attentional copulas
Decoding using attentional copulas

Memory

observed and previously decoded tokens

Sample

Attention

Sample
Decoding using attentional copulas

Memory

observed and previously decoded tokens

Sample

Attention

Memory

observed and previously decoded tokens

Sample

Attention
Decoding using attentional copulas

**Theorem:** decoding in a *random order* guarantees convergence to *valid copulas*
State-of-the-art forecasting performance

CRPS-Sum means (± standard errors). Lower is better. Best results in bold.

<table>
<thead>
<tr>
<th>Model</th>
<th>electricity</th>
<th>fred-md</th>
<th>kdd-cup</th>
<th>solar-10min</th>
<th>traffic</th>
<th>Avg. Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-ARIMA</td>
<td>0.077 ± 0.016</td>
<td>0.043 ± 0.005</td>
<td>0.625 ± 0.066</td>
<td>0.994 ± 0.216</td>
<td>0.222 ± 0.005</td>
<td>4.7 ± 0.3</td>
</tr>
<tr>
<td>ETS</td>
<td>0.059 ± 0.011</td>
<td><strong>0.037 ± 0.010</strong></td>
<td>0.408 ± 0.030</td>
<td>0.678 ± 0.097</td>
<td>0.353 ± 0.011</td>
<td>4.4 ± 0.3</td>
</tr>
<tr>
<td>TempFlow</td>
<td>0.075 ± 0.024</td>
<td>0.095 ± 0.004</td>
<td>0.250 ± 0.010</td>
<td>0.507 ± 0.034</td>
<td>0.242 ± 0.020</td>
<td>3.9 ± 0.2</td>
</tr>
<tr>
<td>TimeGrad</td>
<td>0.067 ± 0.028</td>
<td>0.094 ± 0.030</td>
<td>0.326 ± 0.024</td>
<td>0.540 ± 0.044</td>
<td>0.126 ± 0.019</td>
<td>3.6 ± 0.3</td>
</tr>
<tr>
<td>GPVar</td>
<td>0.035 ± 0.011</td>
<td>0.067 ± 0.008</td>
<td>0.290 ± 0.005</td>
<td><strong>0.254 ± 0.028</strong></td>
<td>0.145 ± 0.010</td>
<td>2.7 ± 0.2</td>
</tr>
<tr>
<td>TACTiS-TT</td>
<td><strong>0.021 ± 0.005</strong></td>
<td>0.042 ± 0.009</td>
<td><strong>0.237 ± 0.013</strong></td>
<td>0.311 ± 0.061</td>
<td><strong>0.071 ± 0.008</strong></td>
<td><strong>1.6 ± 0.2</strong></td>
</tr>
</tbody>
</table>

TACTiS outperforms state-of-the-art models on real-world datasets with hundreds of time series.
State-of-the-art forecasting performance

**TACTiS** outperforms state-of-the-art models on real-world datasets with hundreds of time series.
TACTiS is very flexible

Interpolation

Estimated

5%-95%
10%-90%
25%-75%
50%

Ground truth

Unaligned and non-uniformly sampled data
Thank you!
Please come by our poster!

Code: https://github.com/ServiceNow/TACTiS