Temporal Attention-Gated Model for Robust Sequence Classification

Wenjie Pei¹, Tadas Baltrusaitis², David M.J. Tax¹ and Louis-Philippe Morency²

¹Pattern Recognition Laboratory, Delft University of Technology
²Language Technologies Institute, Carnegie Mellon University

Motivation

Typical sequence classification models are designed for well-segmented sequences, which often need manual and time-consuming processing.

Goal: We propose Temporal Attention-Gated Model (TAGM) to better deal with noisy or unsegmented sequences by automatically locating the salient segments.

Contributions

- Automatically capture salient parts of the input noisy sequence to achieve better performance.
- Inferred attention scores provide meaningful interpretation for the informativeness of each time step.
- Less parameters leading to faster training and better generalizability with less training data.
- Generalization across different tasks and modalities.
- Code available¹.

Model

- **Recurrent Attention-Gated Units**

  To learn an effective hidden representation.

  \[ h_t = (1 - \alpha_t) \cdot h_{t-1} + \alpha_t \cdot h_t' \]

  \[ h_t' = g(W \cdot h_{t-1} + U \cdot x_t + b) \]

- **Temporal Attention Module**

  To extract salient frames.

  \[ \alpha_t = \sigma(m^T (h_t', h_t) + b) \]

  \[ h_t' = g(W x_t + U h_{t-1} + b) \]

  \[ h_t = g(W x_t + U h_{t-1} + b) \]

Experiments

We perform experiments with TAGM on three datasets to show generalization across different tasks and modalities.

- **Speech Recognition**

  Dataset: Noisy Arabic spoken digit dataset (8,000 utterances, 10 digits)

  Feature: MFCCs

  The visualization of attention weights of TAGM on 10 samples (one sample for each digit). For each subfigure, the top subplot shows the spectrogram of the original sequence data while the bottom subplot shows the attention values over time. The red lines indicate the groundtruth of salient segments.

- **Sentiment Analysis**

  Dataset: Stanford Sentiment Treebank (11,855 review sentences, binary-classification or fine-grained task)

  Feature: 300 dimension vectors

  The visualization of attention weights of TAGM on 10 samples (one sample for each digit). For each subfigure, the top subplot shows the spectrogram of the original sequence data while the bottom subplot shows the attention values over time. The red lines indicate the groundtruth of salient segments.

- **Event Recognition**

  Dataset: Columbia Consumer Video Database (9,137 videos, 20 events)

  Feature: CNN features from pre-trained AlexNet model

  The attention weight is indicated for representative frames.

  "TAGM" refers to our model.

References


¹https://github.com/wenjiepei/TAGM.