

Conditional Evidence Stream Generator

Joanna Komorniczak, Tobiasz Puślecki, Paweł Ksieniewicz,
Krzysztof Walkowiak

Certainty-based Domain Selection Framework for TinyML Devices

The streams were synthesized from MNIST and SVHN using an original generator based on the conditional evidence. At the input of the stream synthesis procedure, there is a stationary data set DS . The first processing step is to determine the F factor of its objects, which is a value in the range $0 - 1$, correlated with the difficulty of the sample and determined for each pattern from the DS . To estimate the F factor:

1. Transform DS to its components DS' , using *Principal Component Analysis*, leaving only 80% of the explained variance and standardizing the result.
2. Model a *Gaussian Mixture* for DS' with an assumption of 10 mixture components, assuming that each component has its own single variance.
3. Estimate the density of the *Gaussian Mixture* distribution for each point of DS' . It is important to remember that support is estimated for each component of the mixture.
4. Quantile-normalize the obtained density to a uniform distribution along the object axis – independently in each component.
5. Flatten the obtained representation with the sum of components and perform another quantile normalization to uniform distribution, so that for each point from the original set its mapping to the F factor is obtained.

Having the vector of factors F , it is possible to determine the conditional map CM . It informs generator about the availability of each DS object for each domain building the data stream. Here it is possible to configure the number of domains (m), the number of difficulty oscillation cycles (c) and the thresholding range of the difficulty factor (r). To obtain the conditional map CM :

1. Build a condition basis vector constituting an interval-normalized ($0 - 1$) sampling of the sinusoid at m points in the period from 0 to $2\pi c$. Scale the result to the thresholding range r .

2. Calculate a conditional map CM by comparing the condition basis vector to the vector F , so as to obtain a logical matrix informing whether the F factor of a given object exceeds the domain threshold value.

The final, third component of processing metadata is the domain probability map CP . It informs generator about the probability of selecting an object from a given domain in a given batch of the generated stream. It is calculated according to *instant*, *linear* or *normal* dynamics, in accordance with the standard procedure for generating data streams [?].

To establish a data stream, it is necessary to pass DS , F , CM , and CP to the *ConditionalEvidenceStream* control object. It is responsible for using in each subsequent batch only objects allowed for processing in accordance with the CM conditional map for a specific batch described in the domain probability CP . The code of the generator is publicly available at a GitHub repository¹ and as a *PyPI* package.

¹<https://github.com/w4k2/cesg> – *Conditional Evidence Stream Generator*