**Summary**

**Introduction**

Unsupervised feature learning is important task in machine learning which is facing a lot of challenges. Previous methods in unsupervised feature learning include autoencoders, but faced scalability issues. Better approach was tried by including temporal structure of the data, but the most satisfactory solution is based on probability generative model.

**Aim:**

Design a probabilistic generative model of nonlinear ICA, where observed signals are nonlinear functions of sources and mutually independent and segment wise non stationary.

**Methodology**

* Learn features using non-stationarity of the data (TCL analysis).
* A generative model is proposed with individual components having different distributions in different time windows. We observe nonlinear mixtures of these components.
* Finally, we show that TCL indeed learns nonlinear part of nonlinear ICA leaving the linear mixtures part to be decoded by linear ICA.

**Time Contrastive Learning**

TCL is a multinomial logistic regression which is applied to classify all the data points of a signal into their corresponding time windows. The accuracy of TCL being more than chance level shows us that our network learnt about the structure of the time series in order to classify the data points.

**Procedure**

* Divide into segments indexed by = 1, 2, …. T.
* Associate each data point to its corresponding time window index
* A feature extractor is learnt together with MLR, with a linear regression to classify the data points.

**Claim: Time Contrastive Learning is a probabilistic model**

**Proof:**

According to SoftMax function, we have the probability of the data point belonging to class among 1, 2, … T is given by:

P(

where w1, b1 = 0.

We know from Bayesian conditional probability,

P(.

Assume that we have infinite data and TCL converges to global optimum, then by equating the two conditional probabilities and solving them, we get:

– 3

So, our **intention of generating a probabilistic model** is **satisfied** with the help of TCL.

**Objective 2: Probabilistic Generative Model**

Assume is a smooth invertible non-linear mixture of the source signal s given by where .

**How to define a model for**

Assume that source signals are non-stationary and we propose a generative model for sources based on exponential family.

Non-stationarity is assumed to be slower than the sampling rate. So, when we sample the time windows, we assume that they have a constant distribution within the window i.e stationary within these time windows.

The log-pdf of source signal in time window is defined as

log(

where depends on the time window index ensuring non-stationarity and  is a stationary baseline for all the source signals and is a non-linear function belonging to exponential family.

is a concatenation of all the source signals and V is the number of signals required to construct each source signal.

**Claim: Nonlinear components of the signals are indeed linear mapping of feature extractor**

**Proof:**

**Procedure**

* We generate the source signals given by equation 4 and mixing is done by smooth invertible function f. The f can be realised by passing through a multi-layer perceptron with leaky ReLU’s in middle layers.
* We assume a simple scenario where V = 1, and = 0.
* Let the dimension of the feature extractor h be same as that of x.
* Equating the log-pdf per time window calculated by TCL and data generation i.e (equation 3 and 4), we get:

summation over all the n source signals on RHS as includes all the source signals.

, ) are independent of s. This implies that and must span the same linear subspace and thus can be related by:

Thus, the paper explains that TCL indeed learns the non-linear mapping from to , leaving the linear mixtures part to ICA. The condition is that the source signals must be non-stationary in a suitable way otherwise will be independent of and the source signal will not be used in estimation.