SCATTERING HIDDEN MARKOV TREE

J.B. REGLI, J. D. B. NELSON

UCL, Department of statistical science

ABSTRACT

A Scattering Convolutional Hidden Markov Tree (SCHMT) proposes a new inference mechanism for high-dimensional signals by combining the interesting signal representation created by the Scattering Transform (ST) to a powerful Probabilistic Graphical Model (PGM).
A wavelet scattering network computes a signal translation invariant and stable to deformations representation that still preserves the informative content of the signal. Such properties are acquired by cascading wavelet transform convolutions with nonlinear modulus and averaging operators.
The network's structure and its distributions are described using a Hidden Markov Tree (HMT). This yield a generative model for high-dimensional inference. It offers a mean for performing several inference tasks among which are predictions. The scattering convolutional hidden Markov tree displays promising results on both classification and segmentation tasks of complex images.

***Index Terms*—** Scattering network, Deep network, Hidden Markov Model, Classification

1 Introduction

TBD - 450 words

2 ? ? ?

Motivation and state of the art.
TBD - finish first page + 1/2 column words

3 The Scattering transform

Introduction of SCN
TBD - 1 collumn and a half

4 The Scattering hidden Markov tree :

? introduced the use of scattering networks combined with a support vector machine classifier to achieve competitive

Thanks DSTL/UCL Impact studentship for funding.

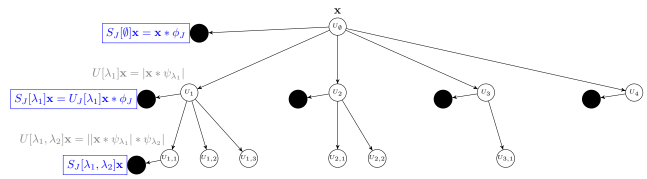

Fig. 1. Frequency decreasing scattering convolution network with $J = 4$, $L = 1$ and $M = 2$. A node i at scale j_i generates $(j_i - 1) \times L$ nodes.

classification performance on some problems. However this method only provides a boolean label for each class. Some methods to express the output of an SVM as a probability exists **?** but they are just a rescaling of the output and not a true probabilistic approach. If one is interested in a true probabilistic model to describe the scattering coefficients, it is quite natural to try expressing them as a probabilistic graphical model. Furthermore generative models are known to be better for inference tasks when the number of training example is low **?**.

4.1 Hidden Markov tree model

We propose an adaptation of those models to create a scattering convolutional hidden Markov tree composed of a set of visible nodes $\{\mathbf{S}_i\}_{i\in\mathcal{T}}$ and a set of hidden node $\{\mathbf{H}_i\}_{i\in\mathcal{T}}$. Both sets are organized in a tree structure such that for any index i of the tree, $S_i \in \mathbb{R}$ and $H_i \in [\![1, K]\!]$ where K is the number of possible hidden states. The initial state is drawn from a discrete non uniform distribution π_0 such that, $\forall k \in [\![1, K]\!]$ $\pi_0(k) = P(H_0 = k)$. For any index i of the tree, the emission distribution describes the probability of the visible node S_i conditional to the hidden state H_i such that, $\forall i \in \mathcal{T}$, $\forall k \in [\![1, K]\!]$ and $\forall s \in \mathbb{R}$ $P(S_i = s_i|H_i = k) = P_{\theta_{k,i}}(s)$, where $P_{\theta_{k,i}}$ belongs to a parametric distribution family and $\theta_{k,i}$ is the vector of emission parameters for the state k and node i. In the remainder of the paper the emission distribution is Gaussian so that $P(S_i = s|H_i = k) = \mathcal{N}(\mu_{k,i}, \sigma_{k,i})$, where $\theta_{k,i} = (\mu_{k,i}, \sigma_{k,i})$ with $\mu_{k,i}$ and $\sigma_{k,i}$ being respectively the mean and the variance of the Gaussian for the k-th value of the mixture and the node i. Finally the probability for the hidden node H_i to be in a state k given its father's state g is characterized by a transition probability such that $\forall i \in \mathcal{T}\backslash\{0\}\ \forall g, k \in [\![1, K]\!]^2\ \epsilon_i^{(gk)} = P(H_i = k | H_{\rho(i)} = g)$ where ϵ_i defines a transition probability matrix such that $P(H_i = k) = \sum_{g=1}^{K} \epsilon_i^{(gk)} P(H_{\rho(i)} = g)$.

Such a model is pictured in Figure 2 and for a given scattering architecture —i.e. fixed M, J and L— the SCHMT model is fully parametrized by,

$$\Theta = \big(\pi_0, \{\epsilon_i, \{\theta_{k,i}\}_{k \in [\![1,K]\!]}\}_{i \in \mathcal{T}}\big). \qquad (1)$$

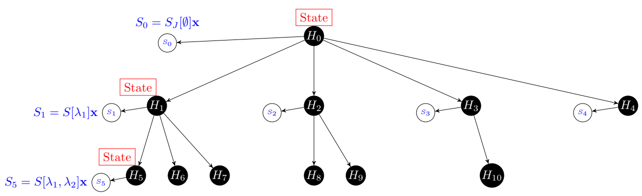

Fig. 2. Scattering convolutional hidden Markov tree.

This model implies to do two assumptions on the scattering transform. First one need to assume — K-populations— that a signal's scattering coefficients can be described by K clusters. This is a common assumptions for standard wavelets [**?**] and hence it can be extended to the scattering transform. The SCHMT also assumed —persistence— that the informative character of a coefficients is propagated across layers. This assumption is sound since ...

4.2 Learning the tree parameters

EM algo

5 Experiments

6 Conclusion

7 COPYRIGHT FORMS

łlabelsec :copyright

You must include your fully completed, signed IEEE copyright release form when form when you submit your paper. We **must** have this form before your paper can be published in the proceedings.

8 REFERENCES