Label Distribution Learning Forests

KAI ZHAO

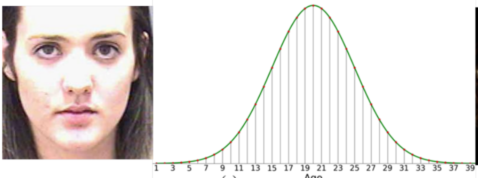
Nankai University

kaiz.xyz@gmail.com

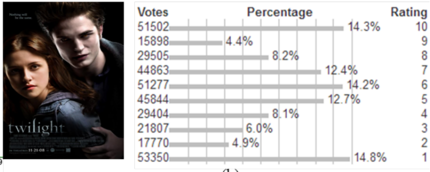
http://kaiz.xyz

November 5

Distribution learning



(a)



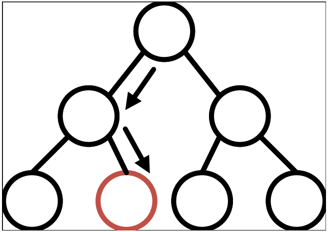
(b)

Example of Distribution Label Annotations:

- (a) Estimated facial ages (a unimodal distribution).
- (b) Rating distribution of crowd opinion on a movie.

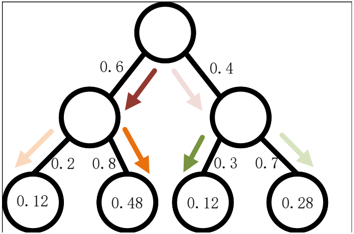
Decision Trees

Discrete Decision Trees:



1. Discrete routing
2. undifferentiable

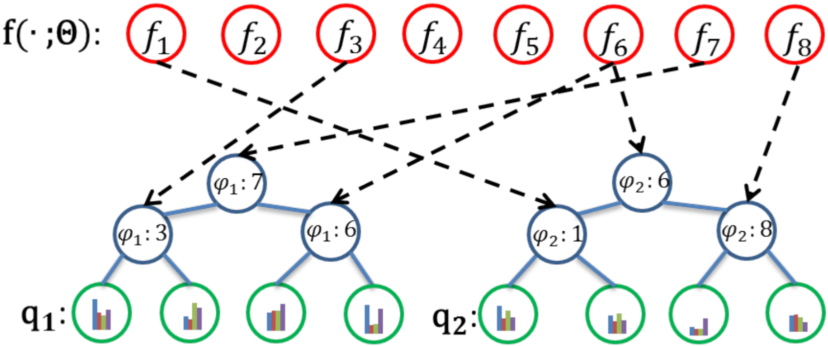
Differentiable Decision Trees:



1. Routing in probability
2. differentiable

$$\mathcal{P}_{\text{leaf}_n} = \prod p_{\text{path}}$$

CNN features as splitting function



- $\mathbf{f}(\cdot, \Theta)$ is hypothesis function of a CNN, and Θ is CNN parameters;
- Randomly select features of from Fully Connected layer;
- Each leaf node contains a distribution;
- Final prediction a a conditional combination of leaf node distributions.

How to Train Proposed Method

The overall parameters of propsoed methods are:

- CNN parameters Θ
- Leaf node distributions $\mathcal{D}$

Data: Initialize Θ and $\mathcal{D}$
while *not convergent* **do**
    while $\mathcal{D}$ *not convergent* **do**
        Training leaf node distribution with EM;
    end
end

Fix Θ to train $\mathcal{D}$ and alternatively fix $\mathcal{D}$ to train Θ

The Loss Function

We choose the widely used *kl-divergence* our loss:

$$l(y, \hat{y}) = -\sum_i \hat{y}_i \log \frac{y_i}{\hat{y}_i}$$

In principle you can choose whatever a **distribution distance** as loss function.

Performance Comparison

Performance of Facial Age Estimation on Morph-II dataset:

Methods	**CPNN**	**BFGS-LDL**	**DLDL+VGG-Face**	**Ours**
MAE	4.870.31	3.940.05	2.420.01	**2.240.02**

References

Shen, Wei and Zhao, Kai and Guo, Yilu and Yuille, Alan (2017)
Label Distribution Learning Forests
Proceedings of Advances in neural information processing systems 2017, USA, Long Beach.

The End

Thank You!

The project page: http://kaiz.xyz/ldlf

Or scan the QR code above.