Dynamics in Classifier

intern

Multimedia Laboratory @ CUHK

Sep 20, 2017

Formulation

Background

- How Imagenet1k is cleaned
  - 671 leaf nodes, 329 internal nodes
  - Classse are balanced
- Difficulty to classify on ImageNet22k (Deng et al. 2009) comes from
  - Fine-grained classed with few instance to train

Hierachical Prediction

Feature of a sample $\mathbf{x}_i$. Ground truth label $\mathbf{y}_i = c_i$, *i.e.* $y_i^{c_i}$.

$$\begin{aligned} P({y_i}^{c_i}|\mathbf{x}_i) &= \Sigma_{c_k \in \textsc{Par}(c_i)} P({y_i}^{c_i}|y_i^{c_k}, \mathbf{x}_i) P(y_i^{c_k}|\mathbf{x}_i) \\ &= P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i)}|\mathbf{x}_i) \end{aligned}$$

Genralize to atribute depth d:

$$P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i,d)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i,d)}|\mathbf{x}_i)$$

Hierachical Prediction

Feature of a sample $\mathbf{x}_i$. Ground truth label $\mathbf{y}_i = c_i$, *i.e.* $y_i^{c_i}$.

$$\begin{aligned} P({y_i}^{c_i}|\mathbf{x}_i) &= \Sigma_{c_k \in \textsc{Par}(c_i)} P({y_i}^{c_i}|y_i^{c_k}, \mathbf{x}_i) P(y_i^{c_k}|\mathbf{x}_i) \\ &= P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i)}|\mathbf{x}_i) \end{aligned}$$

Genralize to atribute depth d:

$$P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i,d)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i,d)}|\mathbf{x}_i)$$

Prediction at depth d

$$P(y_i^{\textsc{Par}(c_i,d)}|\mathbf{x}_i) = \frac{\Sigma_{j \in \textsc{Find}(c_i,0,d)} e^{f_j}}{\Sigma_{j \in All} e^{f_j}}$$

Hierachical Prediction

Feature of a sample $\mathbf{x}_i$. Ground truth label $\mathbf{y}_i = c_i$, *i.e.* $y_i^{c_i}$.

$$\begin{aligned} P({y_i}^{c_i}|\mathbf{x}_i) &= \Sigma_{c_k \in \textsc{Par}(c_i)} P({y_i}^{c_i}|y_i^{c_k}, \mathbf{x}_i) P(y_i^{c_k}|\mathbf{x}_i) \\ &= P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i)}|\mathbf{x}_i) \end{aligned}$$

Genralize to atribute depth d:

$$P({y_i}^{c_i}|y_i^{\textsc{Par}(c_i,d)}, \mathbf{x}_i) P(y_i^{\textsc{Par}(c_i,d)}|\mathbf{x}_i)$$

Prediction conditioned by depth d

$$P(y_i^{\textsc{Par}(c_i,d)}|\mathbf{x}_i) = \frac{e^{f_{c_i}}}{\Sigma_{j \in \textsc{Find}(c_i,d,-1)} e^{f_j}}$$

Hierachical Loss

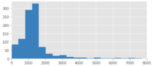
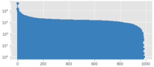
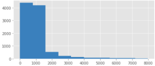
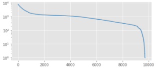
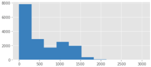
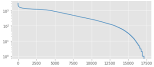
E.g. For Cifar100, given logits $\mathbf{f}$ and ground truth label y_i

$$
\begin{aligned}
L_i^{100} &= -\log\left(\frac{\sum_{j\in \textsc{Find}(y_i,2,2)} e^{f_j}}{\sum_{j\in \textsc{Find}(y_i,0,2)} e^{f_j}}\right)\\
L_i^{20} &= -\log\left(\frac{\sum_{j\in \textsc{Find}(y_i,1,2)} e^{f_j}}{\sum_{j\in \textsc{Find}(y_i,0,2)} e^{f_j}}\right)\\
L_i^{group} &= -\log\left(\frac{\sum_{j\in \textsc{Find}(y_i,2,2)} e^{f_j}}{\sum_{j\in \textsc{Find}(y_i,1,2)} e^{f_j}}\right)
\end{aligned}
$$

Experiments

Statistic Info of ImageNet

Samples/Class

ImgNet	imgs	imgs/cls	histogram	distribution
~1k	~1.4M	~1.4K [1]		
~10k	~28.96M	~2.8K		
~17k	9.4M	545		

[1] images/class of ILSVR2012 competition = 1.3k for all most all classes. We download the latest dataset.

Statistic Info of ImageNet

Hierachical Categories

All sub-datasets are sampled from ImageNet, and 1K covers most of 22k, thus, it is observed that characteristic of categories' distribution are similar.

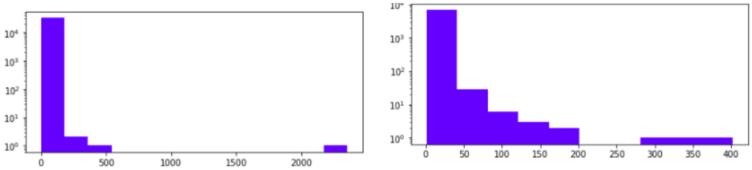


Figure 1: **Left:** Distribution of nodes' child number, **Right**: Igonore outliers

Statistic Info of ImageNet

Hierachical Categories

All sub-datasets are sampled from ImageNet, and 1K covers most of 22k, thus, it is observed that characteristic of categories' distribution are similar.

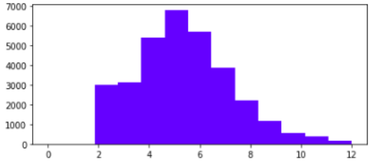


Figure 2: Distribution of nodes' depth

Learning Curve

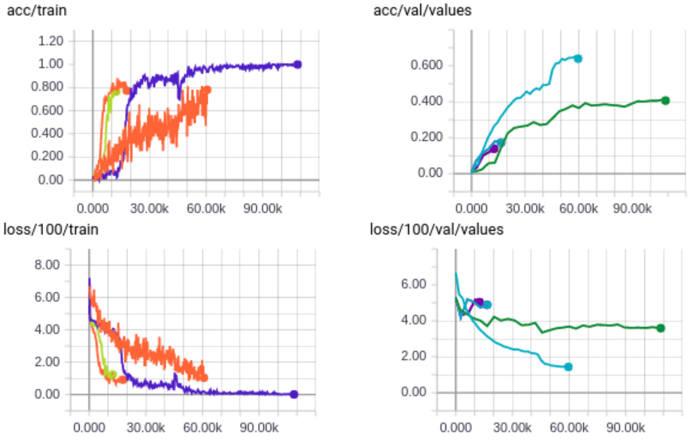


Figure 3

Learning Curve

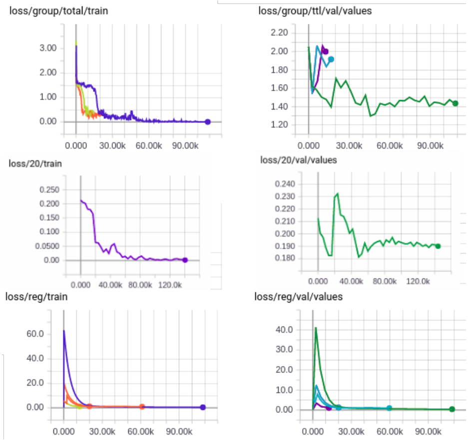


Figure 4: The initial value of loss100 is 5, loss group is 1.6 and loss20 0.27.

Rank

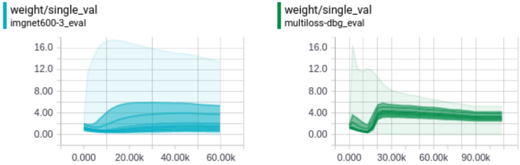


Figure 5: The distribution of weight matrix's sinlge value. It is a common trading on different dataset that the class codes to a concentrating into a subspace

Orthogonality

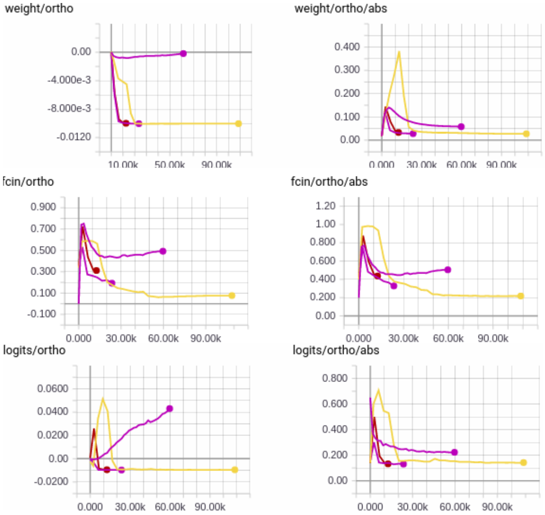


Figure 6: logits=weight^T · fcin. The logits of imagenet become positive may because the reference sample do not cover the whole validation dataset and there are many instance of similar classes.

Reflections

- Generalize to WordNet Tree
  - Key Obstacle: The number of child of node is not same, *i.e.* function FIND cannot feed with *fix-length* matrix and return matrix. Hostile for gpu matrix operation.
- Structure Code for Feature

References I

Deng, Jia et al. (2009). "Imagenet: A large-scale hierarchical image database." In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, pp. 248–255.