

# 北京交通大学

信息网络专题研究课文献阅读报告（应用层—机器学习）



姓名：苏昱臻

学号：17231344

班级：通信 1704 班

学院：电子信息工程学院

任课教师：郭宇春

模块讲授教师：陈一帅

时间：2020 年 5 月 20 日

## 一. 文献信息

文献名称: 《Matching Networks for One Shot Learning》

文献作者: Oriol Vinyals、Charles Blundell、Timothy Lillicrap、Koray Kavukcuoglu、  
Daan Wierstra

发表途径: eprint arXiv:1606.04080

发表时间: June 2016

## 二. 问题意义

### 2.1 研究问题

本文研究的是小样本学习(One Shot Learning), 小样本学习对于传统机器学习来说还是一项重要的挑战。本文作者提出将 metric learning 和 external memories 两种方法相结合, 构造自己的学习网络, 提升小样本学习的准确性。

### 2.2 研究背景

作者指出, 人类在极少的监督下可以学习新的概念, 例如: 一个小孩可以从一本插画的一张图片中学习长颈鹿这个概念; 但对于深度学习来说, 学习一个新的概念需要成千上百次的训练。机器学习需要大量训练的缺陷得到人们的关注, 也推动着小样本学习的研究。作者认为深度学习需要大量训练的原因在于参数模型的特性, 需要大量的训练才可以将参数训练到理想的状态。对于非参数模型来说, 可以快速从新样本中学习到相关特征, 例如 KNN 网络。本文作者要将参数化模型和非参数化模型相融合, 提出自己的模型 Matching Nets。

### 2.3 研究意义

本文研究的小样本学习在保证准确率的情况下, 大大缩短训练模型所花时间。这对于实际应用可以说是革命性的, 可以减少大量训练模型的成本, 提高效率。作者还提出了一种新的训练策略, 在一个训练任务中包含支持集和 Batch 样本。

## 三. 思路方法

### 3.1 模型总体框架

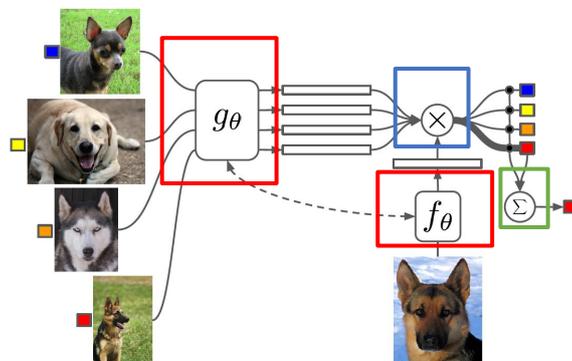


图 3.1-1 Matching Nets 框架

Matching Nets 模型可以由上图直观反映出, 图中红色框线的  $g_{\theta}$  和  $f_{\theta}$  是特征提取函数,

提取训练集和测试样本的特征；蓝色框线为权重判断函数，通过权重判断函数判断测试样本与训练集中各类别的相似程度；绿色框线为加权结果输出函数，将权重判断函数运算的各类别权重与其进行加权求和从而得到测试样本的预测类别。

在作者看来，对于一个给定的训练集 $S$ 和待识别样本 $\hat{x}$ 。作者将 Matching Nets 网络模型数学表述为：

$$\hat{y} = \sum_{i=1}^k a(\hat{x}, x_i) y_i \quad (3.1)$$

上式可以理解为测试样本的预测值是现有训练集中各类别的线性组合。对于 $a$ ，作者认为他是 KDE 算法和 KNN 算法的结合。将 $a$ 看作是 $X \times X$ 矩阵的核，那么 $a$ 就是核密度函数，该模型可以近似为：DL 做嵌入层，KDE 做分类层；如果将 $a$ 看作是与距离度量有关的 0/1 函数，那么该模型就可以认为是：DL 做嵌入层，KNN 做分类层。

### 3.2 The Attention Kernel

作者将 $a(\hat{x}, x_i)$ 看作为 attention kernel。在清楚了 $a(\hat{x}, x_i)$ 的具体作用后，作者进一步介绍 $a(\hat{x}, x_i)$ 的内部机理。 $a(\hat{x}, x_i)$ 函数具体可以分成两步：1) 嵌入函数提取训练集中各类别和测试样本的特征值；2) 计算测试样本特征与各类别特征之间的相似程度归一化权值，这里作者采用的是余弦距离函数和 softmax 计算特征相似归一化权值。 $a(\hat{x}, x_i)$ 的具体实现过程可以由式(3.2)来表示：

$$a(\hat{x}, x_i) = e^{c(f(\hat{x}), g(x_i))} / \sum_{j=1}^k e^{c(f(\hat{x}), g(x_j))} \quad (3.2)$$

其中 $f$ 和 $g$ 是两个嵌入函数，可以由神经网络实现，如在视觉方面的 VGG 或者 Inception。

### 3.3 Full Context Embeddings

在上述基础上，作者认为在提取特征时仅单独考虑 $\hat{x}$ 和 $x_i$ 是短视的，还应该考虑 $\hat{x}$ 、 $x_i$ 和训练集 $S$ 的关系，得到的特征提取器不再是 $f(\hat{x})$ 和 $g(x_i)$ ，而是 $f(\hat{x}, S)$ 和 $g(x_i, S)$ ，在这里作者称之为 FCE(Full Context Embeddings)。作者使用长短期记忆网络 LSTM 得到新的嵌入函数，如下式：

$$f(\hat{x}, S) = \text{attLSTM}(f'(\hat{x}), g(S), K) \quad (3.3)$$

其中 $f'(\hat{x})$ 是 CNN 嵌入层的输出，可以使用 VGG 或 Inception 网络， $g(S)$ 是 support set 中样本的嵌入函数输出， $K$ 是 LSTM 层的训练次数。

## 四. 实验结论

### 4.1 图像分类

作者将自己的模型与现有的 PIXELS、BASELINE CLASSIFIER、MANN、CONVOLUTIONAL SIAMESE NET 四种网络模型进行仿真比较。

#### 4.1.1 Omniglot

Omniglot 是由来自 50 个不同字母表的 1623 个字符组成。由于其具有大量的类别，每个类别的数据又相对较少，使其成为小规模小样本分类的理想数据集。作者构建的训练集为：

独立于字母表，随机选取 $N$ 个字符，生成支持集和 Batch 包进行训练比较。从作者提供的仿真结果可以看出，对于不同的类别数和训练次数，作者所提出的模型均优于其他四种现有模型；作者同时指出，训练次数越多，类别越少，训练得到的结果越精确，这与我们常理所做出的判断是相符合的。

#### 4.1.2 ImageNet

在此部分，作者使用了三个数据集：randImageNet, dogsImageNet 和 miniImageNet。在对 miniImageNet 数据集的仿真比较中，作者指出：其所展示出的结果与在 Omniglot 实验中的结果相同，但与 Omniglot 不同的是，miniImageNet 上使用 FCE 和 fine tuning 会使最终结果的准确度提高 2 个百分点。

在 randImageNet, dogsImageNet 两个实验中，我理解的实验条件是这样的，例如 randImageNet，作者从 ImageNet 随机剔除一些类别 $L_{rand}$ 作为训练集，然后将测试集是否从 $L_{rand}$ 选择作为性能评判条件。可以看出，dogsImageNet 是剔除了一个特定的类别(dog)，对其进行识别更具有挑战性。仿真结果见下表：

Model	Matching Fn	Fine Tune	ImageNet 5-way 1-shot Acc			
			$L_{rand}$	$\neq L_{rand}$	$L_{dogs}$	$\neq L_{dogs}$
PIXELS	Cosine	N	42.0%	42.8%	41.4%	43.0%
INCEPTION CLASSIFIER	Cosine	N	87.6%	92.6%	<b>59.8%</b>	90.0%
<b>MATCHING NETS (OURS)</b>	Cosine (FCE)	N	<b>93.2%</b>	<b>97.0%</b>	58.8%	<b>96.4%</b>
INCEPTION ORACLE	Softmax (Full)	Y (Full)	$\approx 99\%$	$\approx 99\%$	$\approx 99\%$	$\approx 99\%$

从表中可以看出，作者的 Matching Nets 相较于对比模型 PIXELS 和 INCEPTION CLASSIFIER 来说，总体性能是优越的；同时作者向我们介绍了一个 ImageNet 库识别性能特别优越的模型：Inception Oracle。

之后，作者还在语言方面进行了实验，进行缺省语句单词匹配。从作者给出的仿真结果来看，Matching Nets 的准确度并不是很高，没有图像识别那么性能优越，所以在以后的工作中，还需要再语言识别方面进行模型改进。

## 五. 启发思考

本篇论文向我们展示了一种与我们平常所见不同的机器学习方法：小样本学习(One Shot Learning)，结合了参数模型与非参数模型的优点。它不同于传统的深度学习，需要大量的训练次数，少量训练也可以获得较高的识别准确度。我认为小样本学习的特点会使得它在实际中有广泛的应用，可以大大减少训练模型所花费的物力和时间，极大提高使用效率，具有很大的发展前景。

本篇论文偏理论性较多，对多种现有神经网络的优点进行了相结合，对于初识人工智能的我来说具有一定的挑战性，例如 LSTM、KNN 网络，KDE 算法等相关概念。但通过阅读本论文，我也掌握了模式识别的整体思路，那就是首先使用训练集对模型进行训练，提取每个类别的特征，之后对于需要识别的样本，进行特征的提取，然后使用相关数学手段如：余

弦距离，计算样本特征与每个类别特征的相似程度，其中相似程度最大的类别即为模型判断样本所属于的类别。

通过本次论文的阅读，使我对机器学习相关概念有了进一步了解。不是只停留在原来感叹他神奇之处的阶段，而是更多的去了解它背后深层次的运作原理，为我开启了学习机器学习的大门。这对于我进行相关领域的研究和将机器学习的相关方法应用于学习、科研中是有益处的。