介绍

Siamese 网络是一种相似性度量方法，当类别数多，但每个类别的样本数量少的情况下可用于类别的识别、分类等。传统的用于区分的分类方法是需要确切的知道每个样本属于哪个类，需要针对每个样本有确切的标签。而且相对来说标签的数量是不会太多的。当类别数量过多，每个类别的样本数量又相对较少的情况下，这些方法就不那么适用了。其实也很好理解，对于整个数据集来说，我们的数据量是有的，但是对于每个类别来说，可以只有几个样本，那么用分类算法去做的话，由于每个类别的样本太少，我们根本训练不出什么好的结果，所以只能去找个新的方法来对这种数据集进行训练，从而提出了 siamese 网络。siamese 网络从数据中去学习一个相似性度量，用这个学习出来的度量去比较和匹配新的未知类别的样本。这个方法能被应用于那些类别数多或者整个训练样本无法用于之前方法训练的分类问题。

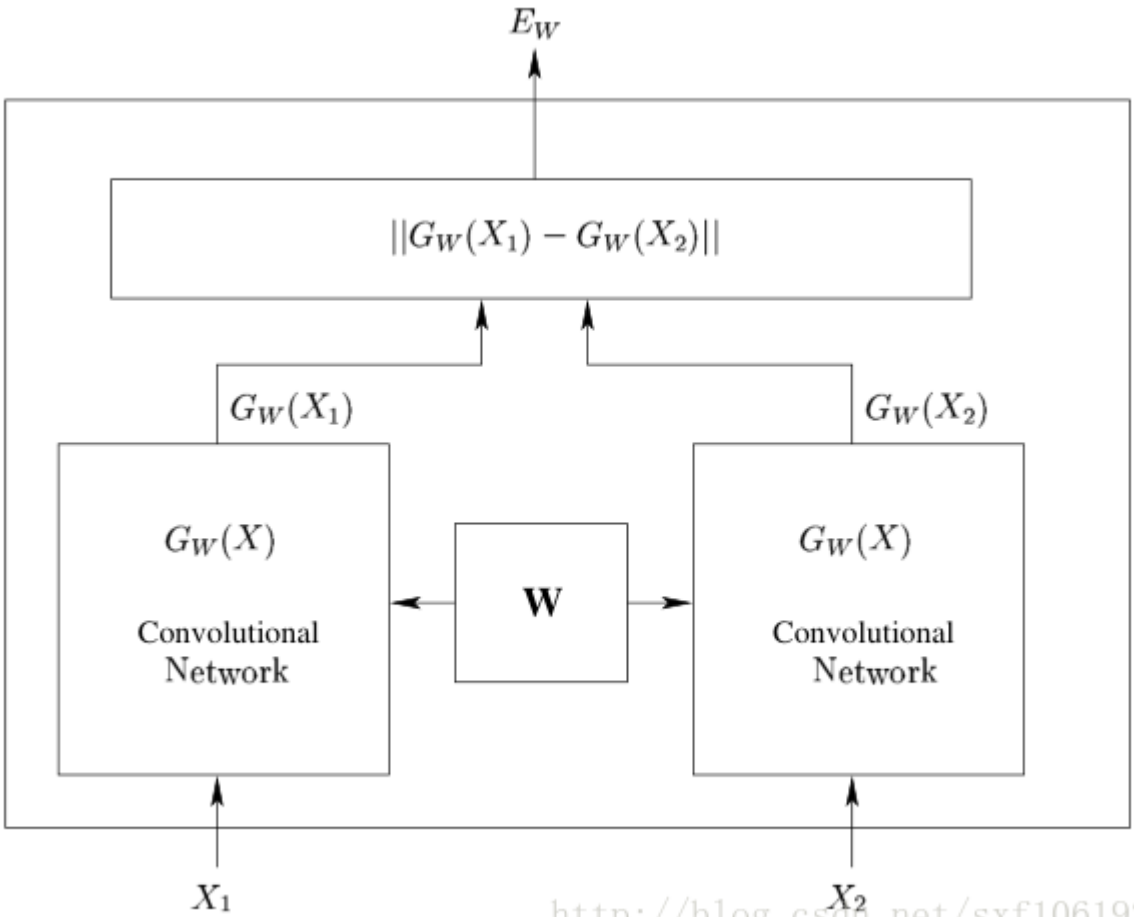
主要思想

主要思想是通过一个函数将输入映射到目标空间，在目标空间使用简单的距离（欧式距离等）进行对比相似度。在训练阶段去最小化来自相同类别的一对样本的损失函数值，最大化来自不同类别的一堆样本的损失函数值。给定一组映射函数 $Gw(X)$,其中参数为 W，我们的目的就是去找一组参数 W。使得当 X_1 和 X_2 属于同一个类别的时候，相似性度量 $E_W(X_1,X_2)=\|G_W(X_1)-G_W(X_2)\|$ 是一个较小的值，当 X_1 和 X_2 属于不同的类别的时候，相似性度量 $E_W(X_1,X_2)=\|G_W(X_1)-G_W(X_2)\|$ 较大。这个系统是用训练集中的成对样本进行训练。当 X_1 和 X_2 来自相同类别的时候，最小化损失函数 $E_W(X_1,X_2)$ ，当 X_1 和 X_2 来自不同类别的时候，最大化 $E_W(X_1,X_2)$ 。这里的 $Gw(X)$ 除了需要可微外不需要任何的前提假设，因为针对成对样本输入，这里两个相同的函数 G，拥有一份相同的参数 W，即这个结构是对称的，我们将它叫做 siamese architecture。

在这篇论文中，作者用这个网络去做面部识别，比较两幅图片是不是同一个人，而且这个网络的一个优势是可以去区分那些新的没有经过训练的类别的样本。

Siamese 也算是降维方法的一种。

网络结构



上图是论文中的网络结构图，左右两边两个网络是完全相同的网络结构，它们**共享相同的权值 W**，输入数据为一对图片（X1,X2,Y）,其中 Y=0 表示 X1 和 X2 属于同一个人的脸，Y=1 则表示不为同一个人。即相同对为（X1,X2,0）,欺骗对为（X1,X2',1）针对两个不同的输入 X1 和 X2，分别输出低维空间结果为$G_W(X_1)$和$G_W(X_2)$，它们是由X_1和X_2经过网络映射得到的。然后将得到的这两个输出结果使用能量函数$E_W(X_1, X_2)$进行比较。

$$E_W(X_1, X_2) = ||G_W(X_1) - G_W(X_2)||$$

损失函数定义

我们假设损失函数只和输入和参数有关，那么我们损失函数的形式为：

$$\mathcal{L}(W) = \sum_{i=1}^{P} L(W, (Y, X_1, X_2)^i)$$

$$L(W, (Y, X_1, X_2)^i) = (1 - Y) L_G \left(E_W(X_1, X_2)^i\right) + Y L_I \left(E_W(X_1, X_2)^i\right)$$

其中 $(Y, X_1, X_2)^i$ 是第 i 个样本，是由一对图片和一个标签组成的，其中 LG 是只计算相同类别对图片的损失函数，LI 是只计算不相同类别对图片的损失函数。P 是训练的样本数。通过这样分开设计，可以达到当我们要最小化损失函数的时候，可以减少相同类别对的能量，增加不相同对的能量。很简单直观的方法是实现这个的话，我们只要将 LG 设计成单调增加，让 LI 单调递减就可以了，但是我们要保证一个前提就是，不相同的图片对距离肯定要比相同图片对的距离小，那么就是要满足：

Condition 1 $\exists m > 0$, *such that* $E_W(X_1, X_2) + m < E_W(X_1, X_2')$,

所以论文中用了一个

$$H(E_W^G, E_W^I) = L_G(E_W^G) + L_I(E_W^I)$$

作为总的损失函数，可以满足这个 condition1。论文中进行了各种假设的证明已经单调性的证明，这里不再重复。

最后给出一个精确的对单个样本的损失函数：

$$L(W,Y,X_1,X_2)=(1-Y)L_G(E_w)+YL_I(E_w)$$

$$=(1-Y)\frac{2}{Q}(E_w)^2+(Y)2Qe^{-\frac{2.77}{Q}E_w}$$

其中

$$E_W = \|G_W(X_1) - G_W(X_2)\|$$

,Q 是一个常量。

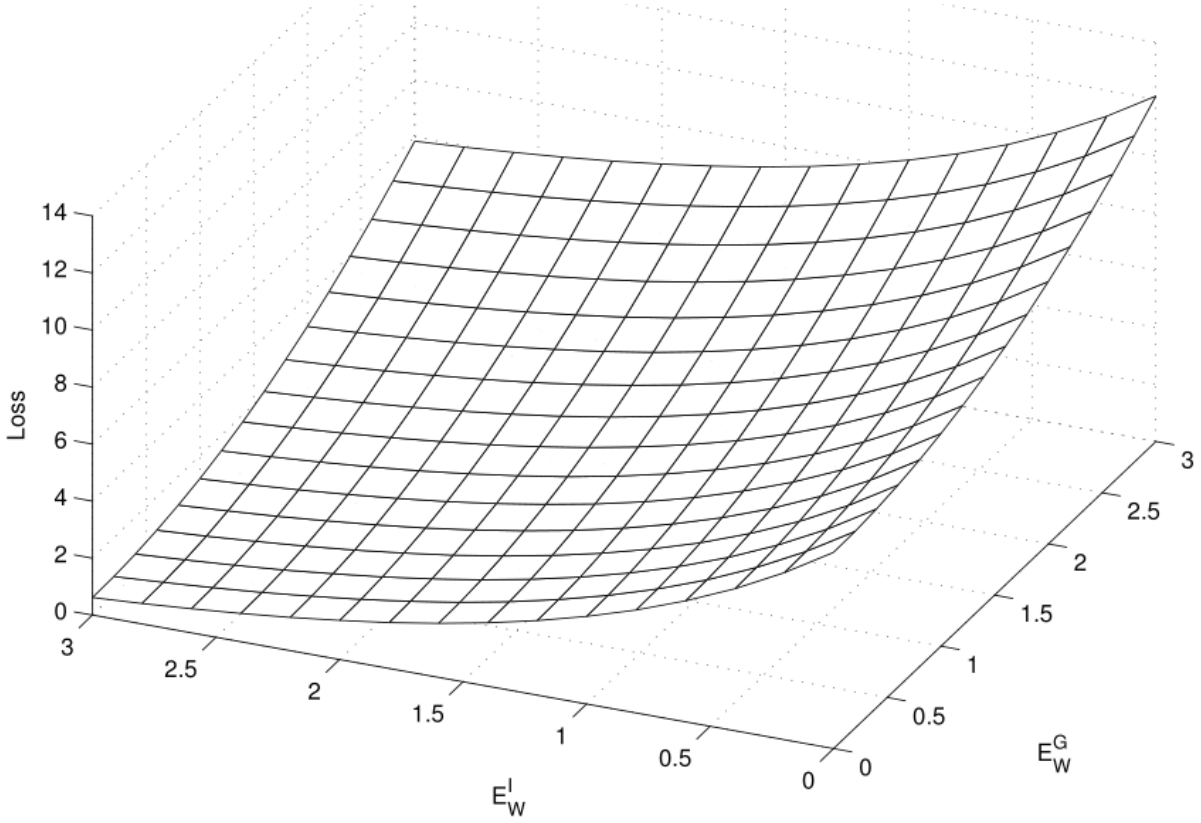


Figure 2. Graph of the loss function H against E_W^G and E_W^I in 3D.

上图说明了收敛性。

总结思想

其实讲了这么多，主要思想就是三点：

1、输入不再是单个样本，而是一对样本，不再给单个的样本确切的标签，而且给定一对样本是否来自同一个类的标签，是就是 0，不是就是 1

2、设计了两个一模一样的网络，网络共享权值 W，对输出进行了距离度量，可以说 l1、l2 等。

3、针对输入的样本对是否来自同一个类别设计了损失函数，损失函数形式有点类似交叉熵损失：

$$L(W,Y,X_1,X_2)=(1-Y)L_G(E_w)+YL_I(E_w)$$

$$=(1-Y)\frac{2}{Q}(E_w)^2+(Y)2Qe^{-\frac{2.77}{Q}E_w}$$

最后使用获得的损失函数，使用梯度反传去更新两个网络共享的权值 W。

优点

这个网络主要的优点是淡化了标签，使得网络具有很好的扩展性，可以对那些没有训练过的类别进行分类，这点是优于很多算法的。而且这个算法对一些小数据量的数据集也适用，变相的增加了整个数据集的大小，使得数据量相对较小的数据集也能用深度网络训练出不错的效果。

实验设计

实验的时候要注意，输入数据最好打乱，由于这样去设计数据集后，相同类的样本对肯定比不相同的样本对数量少，在进行训练的时候最后将两者的数据量设置成相同数量。

总结

本文解释的只是最早提出的 siamese 网络结构，提出的是一种网络结构思想，具体的使用的网络形式完全可以自己定义。包括损失函数，相似度距离的定义等。比如将损失函数的 $2Qe^{-\frac{2.77}{Q}E_w}$ 用 hige loss 代替等。

《Hamming Distance Metric Learning》这篇论文对 siamese 进一步改进，提出了一个 triple net，主要贡献是将成对样本改成了三个样本，输入由（X1,X2,Y）变成了（X1，X2，X1’），表示 X1 和 X1’是相同类别的样本，X1 和 x2 是不同样本的类别。

《Learning to Compare Image Patches via Convolutional Neural Networks》这篇论文写得也很好，将两个网络进行合并，输入的成对标签直接同时输入同一个网络。

代码

使用 tensorflow 在 mnist 上实现的 siamese net

参考文献 2 的官方 code

参考文献

[1] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 539–546. IEEE, 2005.
[2] Mohammad Norouzi, David J. Fleet, Ruslan Salakhutdinov, Hamming Distance Metric Learning, Neural Information Processing Systems (NIPS), 2012.

完