



上海市软件行业协会 组编



2017_年 软件工程论文专集

Proceeding of Software Engineering

上海科学技术出版社





2017年

软件工程论文专集

Proceeding of Software Engineering

上海市软件行业协会 组编

上海科学技术出版社

图书在版编目 (CIP) 数据

2017 年软件工程论文专集 / 上海市软件行业协会组编 .
— 上海 : 上海科学技术出版社 , 2017.11

ISBN 978-7-5478-3697-2

I . ① 2… II . ① 上… III . ① 软件工程 - 文集 IV .

① TP311.5-53

中国版本图书馆 CIP 数据核字 (2017) 第 210830 号

2017 年软件工程论文专集
上海市软件行业协会 组编

上海世纪出版 (集团) 有限公司 出版、发行
上 海 科 学 技 术 出 版 社
(上海钦州南路 71 号 邮政编码 200235 www.sstp.cn)
上海 ***** 有限公司印刷
开本 889×1194 1/16 印张 13
字数 300 千字
2017 年 11 月第 1 版 2017 年 11 月第 1 次印刷
ISBN 978-7-5478-3697-2/TP·54
定价: 48.00 元

编委会名单



主 编

杨根兴

副主编

李 婷

编 委

顾国强 严德铭 姚宝敬

陈春兰 沈 颖 许 静

2017 年软件工程论文专集 || 目录 ||

基于多任务判别特征学习的信用卡识别 何朔.....	1
金融数据中心环境下 SDN 评测标准化进展与解决方案 何朔	6
一种支持大交易量清算系统的研究和设计 吴金坛.....	12
银行卡转接清算系统在“一带一路”国家的建设实践 吴金坛.....	17
浅谈中国软件园区的转型发展之道	
——以上海浦东软件园的实践为例 张素龙.....	21
利用数理统计方法合理预测就餐系统数据 刘梅.....	26
基于图数据库的科技文献知识图谱应用研究 袁汝焱 孙冲霄 戴炳荣 孙旭旦.....	32
一种业务快速恢复的系统及其实施方案 黄爱华 赖志强 蔡夏琳.....	39
针对海量银行卡交易数据流水文件的比对策略及实施方法 李岭.....	44
美国健康医疗典型模式的研究与启示 胡庆林.....	48
互联网域名系统的安全研究 胡庆林.....	52
HMO 模式在医疗费用控制中的分析与研究 章政	58
基于云医疗服务的统一支付平台的设计与实现 冯阳 孟蛟 冯莎莎.....	62
基于云计算的现代医院管理软件研发及应用 周洪.....	68
“互联网 +”时代的金融数据库规范运维 程永新 杨志洪.....	73
敏捷开发项目中评审控制图的运用 沈颖.....	78
对“先看病后付费”服务的调研与分析 冯晓.....	83
商业医疗保险赔付模式的研究 冯晓.....	89
基于声明的身份认证系统的实现 杜红俊.....	95
基于信息权限管理系统的知识管理方案设计 杜红俊.....	102
面向大数据研究、开发和应用的开放服务系统 何东杰.....	107
软件定义存储技术研究与应用分析 何东杰.....	112

吉联云应用平台的研发与实践 成松林 方加来 陈正红	117
基于元数据的多源数据交换的设计及应用 陆奇峰 刘丰源 王凌 旷志光	123
网络信息安全风险及防范措施研究 刘刚 严超 徐绍飞	129
文化创意产业营销服务平台构建研究 屈春光	133
医疗整形机构信息化集成系统构建策略 屈春光	136
基于 ZigBee 网络的智慧校园系统设计与应用 张立为	139
一种安全便捷的互联网有卡支付解决方案 郑建宾	144
一种基于持卡人线下消费行为的推荐方法 郑建宾	150
一种基于软件定义网络的金融系统高可用解决方案 刘国宝	158
一种金融智能 POS 终端的应用管理系统 王琪 何东杰	164
一种新型智能 POS 终端支付系统 王琪 何东杰	169
关于人工神经网络中反向传播算法的研究 余力	174
一种自动识别闲置网络资产统计平台的设计与应用 汪璐	180
基于 Web 的安全渗透及防护研究	
——以某企业门户网站渗透测试为例 张延国 杨亚萍 刘刚	185
企业上云及其评估标准研究 屠娇绒	190
基于 MapReduce 的医疗资源数据归集与应用 张远桂	196

基于多任务判别特征学习的信用卡识别

何朔

(中国银联电子支付研究院 上海 201201)

摘要: 随着国内信用卡市场的迅速发展壮大,各发卡机构推出的信用卡品种日趋增多,信用卡权益和优惠活动纷繁复杂,目前使用卡 BIN 信息无法实现信用卡产品的细分类。近些年,深度学习在计算机视觉任务中取得了瞩目的效果,这也使通过信用卡图片识别的方式实现信用卡产品的细分类成为可能。首先将深度学习的分类模型应用于信用卡识别任务上,通过使用卷积神经网络来提取信用卡的特征,将该特征与模板库中的信用卡特征进行相似度比对得到预测标签。然后针对信用卡的特性进行分析,发现常规的识别模型所存在的几点问题,在此基础上提出基于多任务判别特征学习的信用卡识别方法,此方法在识别率上有显著的提升,并且将比对次数大幅度降低。

关键词: 卷积神经网络,信用卡识别,多任务学习,判别特征

Credit Card Recognition Based on Multi-task Discriminative Feature Learning

HE Shuo

(Research Institute of Electronic Payment, China UnionPay Co., Ltd., Shanghai 201201, China)

Abstract: With the rapid development of the credit card market in China and the increasing of the varieties of credit cards proposed by those credit card issuers, rights and promotions towards credit card become complex enough that it's hardly efficient to accurately classify credit cards based only on BIN information. In recent years, deeplearning has achieved remarkable results in computer vision tasks, which has made it possible to implement credit card classification by those deep methods of image recognition. This paper applies the deep learning model to a credit card recognition task, which is to extract the characteristics of credit card using convolutional neural network, and the label is predicted by the similarity between extracted characteristics and the template database. According to the characteristics of the credit card, we analyze some problems existing in the common conventional neural model, and then propose a credit card identification method based on multi-task discriminative feature learning. In conclusion, the method proves to have significant improvement in recognition accuracy and efficiency of alignment.

Keywords: CNN, Credit Card Identification, Multitask Learning, Discriminative Feature

0 引言

近年来,随着国内信用卡市场的迅速发展壮大,各发卡机构推出的信用卡品种日趋增多,信用卡权益和优惠活动纷繁复杂,用户经常需要查询特定信用卡的权益和优

惠。为了提高广大用户信用卡权益使用的便利性,中国银联积极尝试用户信用卡权益的自动管理功能,即向用户自动提示信用卡附属权益信息,因此信用卡产品细分种类的匹配成为首要解决的问题,而由于卡 BIN 无法实现信用卡产品细分种类的区分,基于图像的识别方法就提供了一种很好的解决途径。

目前卷积神经网络已经在计算机视觉领域的很多任务中取得了非常好的效果,比如图像识别^[1]、物体检测^[2]、图

作者简介: 何朔,男,1978年生,硕士,高级工程师,中国银联电子支付研究院副院长,主要从事及研究领域:云计算、电子支付与电子商务等新技术研究。E-mail: shuohe@unionpay.com。

像分割^[3]等。卷积神经网络端到端的特性使研究者不再需要花费极大的力气提取手工特征。对于图像识别任务,通常有两种做法,对于具有固定数量类别的数据集直接使用网络模型输出预测标签,对于类别过多或者会有新的类别的数据集通常使用网络模型提取特征值,然后用与模板库的相似度比对求得类别标签。

信用卡图片产品分类是一个典型的分类问题,由于信用卡类别比较多,并且随着时间将会有更多类别的卡片发布,要求图像模型具有极强的特征提取能力。为此,我们探索了基于深度学习的信用卡图像分类模型。信用卡产品识别模型首先基于深度卷积神经网络 ResNet,首先训练深度卷积网络模型,然后通过模型提取特征,使用该特征将拍摄图片与模板库图片进行相似度计算得到识别结果,即相似度最高的模板图片作为测试图片的标签。信用卡图像产品分类模型构建过程中,对图片进行切分、特征提取与合并,再进行降维,既提升了信用卡图案特征提取的精度,又保证了特征相似度计算的效率。

通过对信用卡分类问题的进一步分析,对于具有超过 1 000 类的信用卡,使用卷积网络提取的特征计算相似度匹配的方式来预测信用卡的类别具有两个问题:第一是速度慢,第二容易将信用卡识别为不同类银行的卡种类。这两点导致整个识别方案无法在工程中使用。因而本文根据信用卡的特点,提出基于多任务判别特征学习的信用卡识别方法,该方法将模板库按照银行类别分为多个子模板库,通过识别信用卡的银行类别对信用卡进行一次粗分类,然后使用带有判别属性的卷积神经网络提取的特征并与子模板库进行相似度比对,从而得到信用卡的标签。该方法在识别速度与识别率上都实现了进一步提升。

1 基于深度学习的信用卡图片识别

信用卡产品种类识别的目标是实现一个图片匹配模型,用户通过手机拍摄信用卡图片,通过与库中模板卡进行比对,从而识别信用卡的类别。由于信用卡的细分种类与卡号并没有详细的对应关系,因此无法通过卡号判断具体的信用卡产品细分种类。图像识别技术是一种更好的识别途径。

基于深度学习的信用卡产品种类识别模型,首先训练深度卷积网络模型,然后通过模型提取特征,使用该特征将拍摄图片与模板库图片进行相似度计算得到识别结果,

即相似度最高的模板图片作为测试图片的标签。

1.1 信用卡图片训练样本数据集

深度卷积网络模型的训练是一个有监督训练过程,需要带标签的训练数据。由于拍摄图片难以获得,训练数据为 1 200 多张标准模板卡与大量真实卡;测试样本为 100 类实际拍摄信用卡照片,每类包含弱光照、正常光照、较强光照下的三种照片,即 300 张图片。模板库为随机挑选的 300 张样本图片,库里面包含 100 类实际拍摄卡片的模板。

首先使用人工标记的方法给训练样本打标签,对于相同的卡片或者相似的卡片打的标签一致。1 200 多张图片是完全不够用于模型训练的,所以需要进行数据扩充。针对测试样本具有三种光照下的图片,对每张训练样本做如下处理扩充:

(1) 使用公式 $g(x)=\alpha f(x)+\beta$ 对图片像素进行亮度与对比度变换,其中 α 、 β 取不同的值进行对比度与亮度变换。

(2) 对于步骤 (1) 处理后的图片,进行多种小角度变换,通常选择 3° 以内。

(3) 对于步骤 (2) 处理后的图片,加入随机噪声。

通过这三个步骤,每张样本图片生成了 50 多张图片,而且包含了多种亮度与对比度下的模拟图片。

1.2 基于深度卷积网络的信用卡图片分类模型

信用卡图片识别训练模型主要采用 ResNet 残差网络模型,模型结构见表 1。

ResNet 在 ImageNet 数据集上具有很好的效果,使用 ImageNet 训练好的 ResNet 模型在信用卡数据进行微调后将会加速收敛,并且保证较好的识别率。

模型的具体训练过程如下:

(1) 修改模型结构文件中最后一层 fc 层的名字与输出维数即表格中倒数第二行。

(2) 将训练好的 ResNet 模型权值赋值给待训练的模型。

(3) 将图像样本归一化为 224×224 大小,并输入网络,同时将信用卡图片产品种类标注送给 Softmax 分类器;进行一次前向传播,根据 Softmax 输出信用卡图片产品种类判别概率与实际产品种类的差值计算得到 loss。

(4) 使用反向传播算法,根据梯度下降算法沿着 loss 下降的方向微调整个网络的参数值。

(5) 更换图像样本迭代重复上述步骤,直到 loss 下降到

表 1 ResNet 结构

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112 × 112	7 × 7, 64, stride 2				
		3 × 3 max pool, stride 2				
conv2_x	56 × 56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28 × 28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14 × 14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7 × 7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1 × 1	average pool, 1000-d fc, Softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

合适程度后，停止训练。

1.3 信用卡产品种类识别方法

信用卡图片产品种类的识别过程如下：

(1) 将信用卡图片输入到信用卡特征提取模型中，提取信用卡图片的特征向量 B_1 。

(2) 为了充分提取信用卡图片图案中的深层次特征，将信用卡图片平均切为多份，然后对每一份图片分别提取特征向量为 B_2 、 B_3 、 B_4 、 B_5 ……

(3) 将步骤 (1) 和步骤 (2) 的特征向量进行聚合，然后进行 PCA 降维，形成新的特征向量 B 。

(4) 计算特征向量 B 与特征库存放向量的相似度，譬如使用余弦相似度：

$$\cos \theta = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

找到余弦值最大的特征向量对应的信用卡产品种类，即为待识别信用卡图片的产品种类。

2 基于多任务判别特征学习的信用卡识别

2.1 粗分类标签

使用 ResNet 提取的特征进行 1 : N 相似度比对的方式来识别信用卡，整体所花时间较长。对于误分类的图片

进行查看，发现有很大部分错误标签与真实标签不属于同一银行。该部分信用卡图案的特点是图案的纹理较浅（图 1），相对文字与 Logo 区域的纹理更明显。对于较浅的纹理图案，信噪比低，难以通过深度卷积网络提取到有效信息，这导致直接训练得到的网络模型的识别率不高。



图 1 信用卡图片

如果卷积网络能够从信用卡的文字、Logo 区域提取出信用卡的所属银行类别信息，我们对信用卡模板库按照银行信息分割为多个子数据库，通过所识别的银行类别信息将信用卡归类到所属大类，然后再通过相似度计算在子数据库中进行比对，这对于通过计算相似度的 1 : N 比对系统来讲，将会提高准确度与比对速度。

2.2 判别特征学习

对于训练样本的类别包含了所有可能测试样本的类别（即不会增加新的类别）的识别任务，深度学习的分类模

型能够很好地直接解决该类任务。深度卷积网络将原始图片数据映射为高维向量, 然后通过全连接层进行分类预测标签, 最后的全连接层类似于一个线性分类器, 将高维向量样本进行分离。因而使用 Softmax loss 的卷积网络所学到的特征具有可分离性。Softmax loss 公式如式 (2) 所示:

$$L_S = -\sum_{i=1}^m \log \frac{e^{W_{ji}^T x_i + b_{ji}}}{\sum_{i=1}^n e^{W_{ji}^T x_i + b_{ji}}} \quad (2)$$

式中, $x_i \in R^d$ 表示属于第 i 个样本的深度特征, 它的类别为第 y_i 类, d 为特征维度; $W_j \in R^d$ 表示全连接层 $W \in R^{d \times n}$ 权值的第 j 列; $b \in R^n$ 表示偏移量。训练时每批次输入的样本数为 m , 样本类别数为 n 。

对于信用卡分类, 由于信用卡新卡将会持续发放, 识别的类别将会随时间增多, 这就要求卷积网络学习到的特征不仅需要可分离性, 还需要判别性, 对于新类别标签图片具有通用性。特征的判别性则要求样本的特征分布为类内紧凑和类间可分。

为了提高卷积网络提取的特征更具有判别性, 通常可以对最后的 loss 加一些判别方法的惩罚项, 这里从最小化类内距离与最大化类间距离两个方向进行添加。

类内距离的 loss 项即 centerloss^[4], 公式如式 (3) 所示:

$$L_C = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (3)$$

式中, $c_{y_i} \in R^d$ 表示第 y_i 类深度学习特征的中心。

类间距离的 loss 项公式如式 (4) 所示:

$$L_W = \frac{1}{2} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \|c_i - c_j\|_2^2 \quad (4)$$

目标是最小化类内距离与最大化类间距离, 将 L_C 与 L_W 结合得到判别特征 loss, 如式 (5) 所示:

$$L_D = L_C - L_W \quad (5)$$

对于每次迭代, 需要计算 L_D 关于 x_i 的梯度用于计算卷积网络的反向梯度用于更新权值, 同样需要更新 c_{y_i} , 分别如式 (6)、式 (7) 所示。在实际应用中为了提高算法的鲁棒性, 使用参数 α 控制类别中心的学习率:

$$\frac{\partial L_D}{\partial x_i} = x_i - c_{y_i} \quad (6)$$

$$\Delta C_j = \frac{\sum_{i=1}^m \delta(y_i = j) \cdot (c_j - x_i)}{1 + \sum_{i=1}^m \delta(y_i = j)} - \sum_{i=1}^m (c_i - c_j) \quad (7)$$

这里 $\delta(y_i = j)$ 为在 $y_i = j$ 时的值为 1, 其他情况为 0。联合 Softmax loss 与判别特征 loss 得到新的 loss 公式如式 (8) 所示:

$$\begin{aligned} L_2 &= L_S + \lambda L_D \\ &= -\sum_{i=1}^m \log \frac{e^{W_{ji}^T x_i + b_{ji}}}{\sum_{i=1}^n e^{W_{ji}^T x_i + b_{ji}}} + \\ &\quad \frac{\lambda}{2} \left(\sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 - \sum_{i=1}^{m-1} \sum_{j=i+1}^m \|c_i - c_j\|_2^2 \right) \end{aligned} \quad (8)$$

式中, λ 为平衡两个损失函数的权值参数。

2.3 多任务判别特征学习

由于信用卡银行属性与信用卡类别两个识别任务都可以使用卷积神经网络来完成, 可以将两个任务共用一个卷积网络, 即采用多任务深度学习, 其模型如图 2 所示。

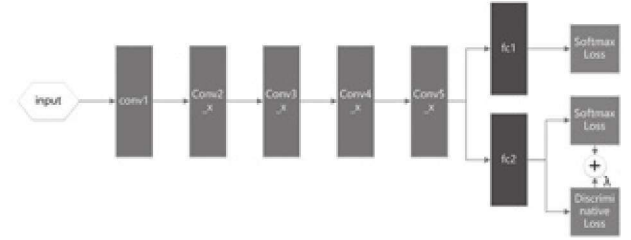


图 2 多任务判别特征学习网络结构

任务一为信用卡银行属性识别, 即图中的 fc1 全连接层后接 Softmax loss, 该层的输出维度为信用卡银行的数目。任务二用于识别信用卡的种类, 即图中的 fc2 全连接层后接 Softmax loss 与判别特征 loss。最后的损失函数为两个任务的损失函数的加权之和, 如式 (9) 所示:

$$L = (1 - \beta)L_1 + \beta L_2 \quad (9)$$

式中, L_1 的计算公式为式 (10):

$$L_1 = -\sum_{i=1}^m \log \frac{e^{W_{ji}^T x_i + b_{ji}}}{\sum_{j=1}^h e^{W_{ji}^T x_i + b_{ji}}} \quad (10)$$

式中, h 为银行类别数目。

整个模型的学习算法如下:

算法: 多任务判别特征学习算法。

输入: 训练集 $\{x_i, y_i, z_i\}$, 其中 x_i 为图像, y_i 为信用卡的类别, z_i 为信用卡的银行类别。初始化卷积层参数 θ_C , 任务一的 loss 层参数 W_1 , 任务二的 loss 层参数 W_2 , 信用

卡类别中心值 $\{c_j | j=1,2,\dots,n\}$ 。超参数 α 、 β 、 λ ，迭代数 $t=0$ ，学习率 μ^t 。

输出： θ_c 、 W_1 。

(1) 从训练集选取 m 个数据， $t=t+1$ 。

(2) 计算联合损失函数值 $L^t = L_1^t + L_2^t$ 。

(3) 计算反向传播误差 $\frac{\partial L^t}{\partial x_i^t} = \frac{\partial L_1^t}{\partial x_i^t} + \alpha \frac{\partial L_2^t}{\partial x_i^t}$ 。

(4) 更新参数 $W_1^{t+1} = W_1^t - \mu^t \cdot (1 - \beta) \frac{\partial L_1^t}{\partial W_1^t}$ 。

(5) 更新参数 $W_2^{t+1} = W_2^t - \mu^t \cdot \beta \cdot \frac{\partial L_2^t}{\partial W_2^t}$
 $= W_2^t - \mu^t \cdot \beta \cdot \frac{\partial L_s^t}{\partial W_2^t}$ 。

(6) 更新每一类信用卡特征中心值 $c_j^{t+1} = c_j^t - \alpha \cdot \Delta c_j^t$ 。

(7) 更新参数 $\theta_c^{t+1} = \theta_c^t - \mu^t \sum_{i=1}^m \frac{\partial L^t}{\partial x_i^t} \cdot \frac{\partial x_i^t}{\partial \theta_c^t}$ 。

(8) 判断是否满足停止迭代条件，是则停止，否则更新 $t=t+1$ 并跳转到 (2)。

2.4 基于多任务判别特征学习的信用卡识别

通过 2.3 节的训练得到的深度卷积神经网络模型在验证集上对于任务一即信用卡的银行类别识别有很高的识别率，对于信用卡类别识别的准确率也有一定的提升。

使用该模型进行信用卡识别，其简要流程如下：

(1) 将信用卡图片输入到卷积神经网络中，得到任务一的 Softmax 结果，即银行类别的预测标签 z_i ，并提取信用卡图片的特征向量 B 。

(2) 利用银行类别标签 z_i 定位到相应类别的子模板库。

(3) 计算特征向量 B 与子模板库的特征库存放向量的相似度，找到相似度最大的特征向量对应的信用卡产品种类，即为待识别信用卡图片的产品种类。

3 实验分析

首先对信用卡训练样本集合与银行类别进行统计，由于银行数目较多，而且很多小银行发行的信用卡类别较少，因而按照银行发行卡的数目保留前 15 类银行，并将其其他银行归并到这 15 类银行中，使用该 15 类银行对所有样

本进行打标签。使用多任务特征学习算法对数据集进行训练，这里取参数 $\alpha=0.6$ 、 $\beta=0.5$ 、 $\lambda=0.002$ ，在验证集上得到的结果见表 2。本文方法增加了信用卡银行类别识别任务与对信用卡识别任务增加判别 loss 后，信用卡误识别为其他银行卡类的比例降低了 5%，而信用卡识别准确率 Top1 提高了 4%，Top5 提高了 6%。

表 2 实验结果对比

方法	银行类别识别率	信用卡准确识别率	
		Top1	Top5
ResNet	91%	86%	92%
ResNet+多任务 +Discriminant loss	96%	90%	98%

在实际应用中，使用该模型得到信用卡的银行类别信息与卷积网络特征在每张信用卡特征比对次数上降低到原来方法的 1/15，这使识别速度有了很大的提高。同时，识别准确率也得到很大的提升。

4 结语

本文针对信用卡图片文字图案纹理明显与信用卡模板库可分为多个子模板库的特点，提出基于多任务判别特征学习的信用卡识别方法，在最终识别率与速度上都实现了很大的提升，使信用卡图片识别技术可用于产品业务之中。

参考文献

- [1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. Doha: International Conference on Neural Information Processing Systems. Curran Associates Inc, 2012: 1097.
- [2] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137.
- [3] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Boston: Computer Vision and Pattern Recognition. IEEE, 2015: 3431.
- [4] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[J]. Springer International Publishing, 2016, 47(9): 499.

金融数据中心环境下 SDN 评测标准化进展与解决方案

何朔

(中国银联电子支付研究院 上海 201201)

摘要: 中国银联自 2014 年开始关注和跟踪软件定义网络 (SDN) 技术的发展,并在三年多的时间里做了大量的研究和探索工作,对 SDN 本身以及在金融行业中的应用前景有着自身的理解。根据金融机构在 SDN 产品测试选型中产生的实际需求,中国银联联合产业各方开展了金融数据中心 SDN 评测标准化的工作,旨在推动 SDN 在金融行业的落地。以此为基础,对标准中评测模型的建立、评测指标的选取以及基于权重的评分方法进行了详细的介绍,并规划了下一步的工作。

关键词: 金融行业, 软件定义网络, SDN, 评测标准

Progress and Solution of SDN Evaluation's Standardization in Financial Data Center Environment

HE Shuo

(Research Institute of Electronic Payment, China UnionPay Co., Ltd., Shanghai 201201, China)

Abstract: China UnionPay has been focusing on and tracking the development of SDN technology since 2014 and has done a lot of research and exploration work in more than three years. China UnionPay has our own understanding of SDN technology and application in the financial industry. According to the actual needs of financial institutions in SDN product testing and selection, China UnionPay has jointed all the industry of SDN and carried out SDN evaluation standardization work of financial data center, which is aimed to promote financial industry implement SDN technology successfully. Based on that, this paper introduces the establishment of evaluation model, the selection of evaluation index and the valuing method based on weight, and releases the plans of the next step.

Keywords: Financial Industry, SDN, Evaluation Standardization

0 引言

近年来,云计算、大数据、区块链等创新技术在数据中心的应用,催生出了新的更复杂的数据网络模型;此外,数据中心的规模还在持续快速增长,为了获得更灵活的网络资源调度能力和更高效的运维能力,也对数据中心的网络环境提出了更高的技术要求。数据中心的网络架构的升级已成为大势所趋,软件定义网络 (SDN) 作为未来网络

的发展方向,已受到越来越多的关注^[1]。

就金融行业来说,随着金融和互联网不断深度融合,金融业务在不断丰富的同时也对网络服务提出了新的挑战。首先,面向互联网方式的金融创新应用的快速发展对网络服务提出更敏捷的服务要求,网络资源的开通与变更从原本的以周为单位提升到分钟级别,以支撑应用快速投产;其次,应用的多样性也导致数据流量结构更为复杂,如何保证对请求的快速响应,为客户提供更流畅的产品体验也是一个难点。为了应对这些挑战,金融行业也开始逐步探索 SDN 技术,从而更有力地推动金融 IT 向高效服务化转型。

在新技术和业务的双因素驱动下,金融行业未来将会

资助项目: 国家云计算示范工程项目 (C7362398902020110006)、上海市青年科技英才扬帆计划 (17YF1425800)。

作者简介: 何朔,男,1978 年生,硕士,高级工程师,中国银联电子支付研究院副院长,主要从事及研究领域: 云计算、电子支付与电子商务等新技术研究。E-mail: shuohe@unionpay.com。

持续加深对 SDN 的研究和应用。在此过程中, 基于金融特色的 SDN 技术体系会逐渐成型, SDN 技术在金融行业内的标准化也会成为必然的趋势。基于此, 中国银联针对金融机构在 SDN 技术选型中产生的评测需求, 展开金融 SDN 评测标准化的研究。

1 SDN 基本概述

1.1 SDN 定义与架构

SDN 是当前最热门的网络技术之一, 它解放了手工操作, 减少了配置错误, 易于统一快速部署, 是业界一致认为解决上述核心问题的网络关键技术^[2]。

作为一整套的网络解决方案, SDN 包含了许多组件和设备。根据业内比较主流的观点, 将 SDN 网络分为四个部分: SDN 控制器、物理交换网络、虚拟交换网络和网络功能虚拟化(图 1)。

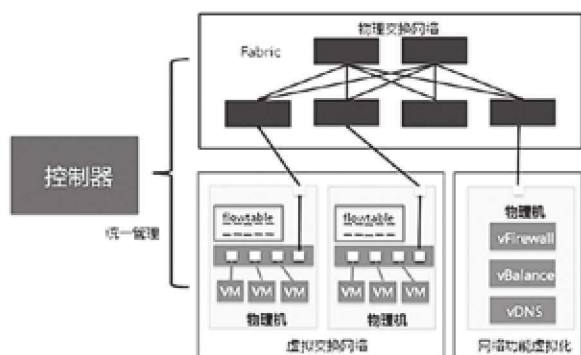


图 1 SDN 各组件结构

1.2 SDN 在金融数据中心内部的应用价值

在金融行业处于业务变革优化的关键时期, 金融数据中心面临着以云计算为代表的新技术的应用、业务的快速上线与变更以及运维自动化这三大挑战, 数据中心内部网络技术的变革是创新已成必然趋势^[3]。而 SDN 被引入数据中心应用后, 其相关特性可以有效解决上述问题。

(1) SDN 基于网络虚拟化技术, 可以实现网络资源的共享与对虚拟机等计算资源动态迁移感知, 该特性可以为云计算网络多租户与虚拟机迁移的能力提供良好支撑^[4]。

(2) SDN 的网络可编程性思想将网络资源服务化, 为上层的业务应用提供了开放性、标准化的接口, 上层应用可根据需求对网络进行灵活调整, 达到了对业务网络需求

分钟级的响应能力。此外, SDN 还具备对网络性能进行优化的能力, Google 在之前就公布了他们利用 SDN 技术把数据中心间的核心网络带宽利用率提高到了 100%, 在降低成本的同时也提高了用户体验。

(3) SDN 将网络控制平面上收, 在控制器中可自动生成全网拓扑, 所以基于此 SDN 可实现全网拓扑变化的集中感知, 从而快速故障定位, 并通过智能决策分析实现故障网络路径规避, 大幅度地提升了数据中心网络运维能力^[5-6]。

2 SDN 在金融行业的应用现状

软件定义网络技术在金融机构已普遍受到重视, SDN 的关键技术特性, 如网络多租户、精细化流量调度、高效率运维已被行业普遍认同, 并被认为是发展方向^[7]。

从应用情况来看, 国内大型金融机构已纷纷开展相关技术的研究与应用探索, 但普遍处于在研发测试环境中探索与尝试的阶段。如工商银行、平安银行、招商银行等机构已经开始与厂商合作, 在研发测试环境进行 SDN 技术应用尝试, 后续也准备在生产环境落地^[8]。

但是当前在应用 SDN 的过程中也出现了一些问题, 而最不可避免的一个就是面向金融行业的 SDN 产品与解决方案参差不齐, 容易因软件和设备能力不足导致基础设施运行稳定风险, 造成生产事故。这是以安全和稳定当先的金融行业不得不重视的问题, 也是 SDN 迟迟不能在金融生产核心落地的根本原因。

3 金融行业对 SDN 技术应用的标准化测试需求

从长远来看, 金融行业必然是 SDN 技术和产品的关键应用方。对于一般金融机构实施过程中, 最主要关心的是设备解决方案的选型问题。而一套富有行业特色的标准化的 SDN 评测方案就成为各金融机构的关注点。主要原因有以下几点:

(1) 金融机构希望通过行业统一的 SDN 测试标准, 规范行业内 SDN 产品质量。

(2) 标准具有全面的框架性评测体系, 能在金融机构进行 SDN 产品选型评测中给予指导。

(3) 标准能够统一金融行业的特色需求, 更方便 SDN 厂商针对金融特性进行产品完善。

(4) 标准能够增强金融机构对 SDN 技术的掌控能力,在一定程度上能够摆脱 SDN 厂商的绑定,提高自身的话语权。

4 金融行业数据中心 SDN 评测标准化发展

4.1 金融 SDN 评测标准化工作里程碑

中国银联在 2016 年初就开始规划金融数据中心 SDN 评测的标准化工作。当年 7 月,中国银联联合复旦大学、国家宽带网络研究中心、上海银行、思博伦,成立了金融数据中心 SDN 评测标准化工作组,正式开始 SDN 评测标准的制定工作。

同年 12 月,《金融数据中心 SDN 评测标准》(V1.0 版本)正式成稿并对外发布。在标准制定过程中,工作组规模不断壮大。截至初稿发布,参与标准制定工作的单位已增加到了 18 家,华为、思科、华三、博科等核心网络设备企业也已加入联合工作组。

2017 年 3 月,该标准在中国通信标准化协会正式立项,标志着金融 SDN 标准化工作已得到业内权威机构的认可。

SDN 评测标准化工作里程碑如图 2 所示。



图 2 SDN 评测标准化工作里程碑

4.2 金融行业数据中心 SDN 评测模型

SDN 作为一种新的网络架构,在对其面向金融行业进行标准化评测时,首先必定要建立新的评测模型。评测模型建立分为以下两个步骤。

4.2.1 步骤 1: 新网络架构下的新评测体系, SDN 评测体系从扁平化向层次化演进

传统网络是由交换机、路由器组成,复杂一些的情况只是再加上防火墙、负载均衡等设备,各个设备的功能特性在整体网络的能力上都能有所体现,所以在传统网络的评测中,只对整体网络进行评测已经足够^[9]。

但是与传统网络功能扁平化的特点相比,SDN 网络则更具备层次性。这种层次性可以从以下两个方面来看:一是在组网结构上,传统网络还只是停留在硬件设备上,而 SDN 已经渗透到了物理机内部的软件。二是从承载的业务流量上看,传统网络中不管是虚拟机还是物理机的流量,都是在同一个层面上进行转发,而 SDN 网络中的虚拟机和物理机的网络平面是分开的,物理交换网络和虚拟交换网络相互配合来实现业务数据的转发。

由此可以看出,SDN 网络的能力特性也必然是层次化的,某个组件的能力并不能纯粹地体现到整网层面上。如果仍然像传统网络评测一样,只对 SDN 整体网络进行评测的话,得到的一些评测结果很可能会掩盖组件功能上的一些缺陷,在实际应用中往往就会出现“木桶效应”。此外,如果对各组件情况不了解,那么在后期与应用平台进行对接的过程中也会遇到很大的困难。

因此,为了保证评测标准的准确性,我们决定从 SDN 整网和组件两个层次上对 SDN 建立评测框架,共包括了五个方面。评测框架如图 3 所示。

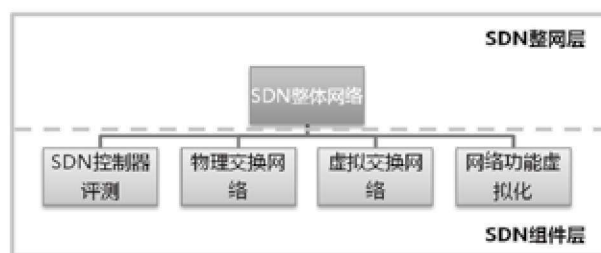


图 3 金融数据中心 SDN 评测框架

4.2.2 步骤 2: 在 SDN 网络本质的基础上整合金融行业和 SDN 自有特色,拓宽评测维度

评测维度的选取则要考虑从哪些方面对 SDN 进行评估。一个合格的评测维度必须恰好涵盖到金融机构在应用 SDN 中所关注的点,既不缺少也不冗余。

为了使金融 SDN 评测维度能够更权威和全面,我们对评测维度的选定步骤也进行了规划,总结下来按照以下三个子步骤来进行选取。

4.2.2.1 子步骤 1: 继承传统网络评测维度

SDN 本质还是一种网络解决方案,所以在对网络能力、

网络技术、网络设备等相关方面的国际和国内标准进行梳理和分析后，选定了 SDN 网络和各个组件所具有的最根本

的四个评测维度。具体见表 1。

表 1 金融数据中心 SDN 基本评测维度

评测维度	选择依据
功能性	一个成熟的网络方案应该具有完备的功能性，满足用户各种明确和潜在的需求，这是网络最基本的特性
性能	性能也属于网络的基本特性，金融行业的业务系统访问者多且流量较大，并且在节假日或者活动期间会出现访问量激增的现象，所以金融行业数据中心网络的性能应具有较高的水平
可靠性	金融行业应用系统往往需要 7×24 h 正常运行，一旦网络中断会造成不利的社会影响，所以网络的可靠性十分重要
可维护性	金融数据中心在规模日益扩大、业务关系不断复杂化的环境下，网络的可运维性也显得愈发重要

4.2.2.2 子步骤 2：结合 SDN 自身特性

SDN 与传统网络在技术、架构等方面存在着许多不同之处。在这里通过对 SDN 三个特性的分析，拓宽评测维度。

1) 控制平面集中管理

逻辑上集中的控制能够支持获得网络资源的全局信息，并根据业务需求进行资源的全局调配和优化。该特性可融入网络的功能性当中体现。

2) 开放接口

通过开放的南向和北向接口，能够实现应用和网络

的无缝集成，使得应用可以根据自己的需求去调整网络。

SDN 这个特性是基本评测维度中没有包括的。但是如果将开放接口单独作为一个维度则略显单薄，与其他维度无法对等，所以将该特性暂且搁置。

3) 网络虚拟化

SDN 屏蔽了底层物理转发设备的差异，实现了底层网络对上层应用的透明化。逻辑网络和物理网络分离后，逻辑网络可以根据业务需要进行配置、迁移，不再受具体设备物理位置的限制。该特性可融入功能性维度中。

4.2.2.3 子步骤 3：融入金融行业特色需求

不同的应用场景会产生出不同的需求，所以对金融行业 SDN 进行评测就必须着重考虑金融行业需求的特殊性。

1) 高安全性

金融数据高度敏感，所以要求网络必须能够保证数据的完整性和秘密性，不仅对外部的恶意攻击具有很强的抵御能力，对内也要满足很高的合规性要求，对任何访问与操作有着严格的控制和审计^[10]。金融监管机构也对金融数据中心网络的安全等级提出了明确要求。在这里将安全性

作为一个评测维度。

2) 多中心的灾备模式

金融行业特有的“两地三中心”的灾备模式，对网络跨数据中心通信的能力提出更高的要求。该特性可融入可靠性中。

3) 网络可扩展

近年来金融业务发展迅速，且各种前沿技术如云计算、大数据也在金融机构落地应用。所以数据中心网络必须能对规模的不断扩张进行支撑，并且有着与云计算、大数据等平台良好对接的能力，从而满足金融数据中心未来的发展需求。

可以看出，该特性在一定程度上包括了 SDN 开放接口特性的内容。所以可以将开放接口特性与之结合，形成一个新的维度——可扩展性。

至此，金融数据中心 SDN 评测模型已建立完成，模型从两个层次、五个方面、六个维度上对金融数据中心 SDN 实施评测。

4.3 金融行业数据中心 SDN 评测指标

评测指标是开展实际评测工作的基础。根据具体的评测指标，测试人员才能制定相应的测试案例和测试方法。评测指标选取的是否合适、得当，对整个评估结果的准确性、公正性会造成非常大的影响。因此工作组讨论制定了金融数据中心 SDN 评测指标的选取所遵循的原则和标准。

4.3.1 金融数据中心 SDN 评测指标选取原则

(1) 典型性。在评估中选取的评测指标要能够充分代表被测对象的某项成熟度特性。

(2) 可测性。在评估中选取的评测指标必须具有可评测

性,也就是说评测指标可以通过工具测试、数据统计、数据调查等方法取得评测结果。

(3) 简明性。评测指标的选取必须在具有充分代表性的基础上力求评测指标含义的简明化,不能出现歧义。

(4) 完备性。完备性是对所有评测指标整体而言的。在评测中定义的所有评测指标要能够涵盖被测对象的基本特征。

(5) 客观性。在评测指标定义中,评测指标本身只反映软件的某种特性,不可带有任何倾向性。

4.3.2 金融数据中心 SDN 评测指标选取标准

(1) 评测指标应该紧紧围绕着金融行业的特点来进行选取和定义。

(2) 评测指标要全面地反映出 SDN 网络或相关组件的特性,且各个评测指标在影响力的层级上应是对等的。

(3) 各评测指标之间应该没有关联或者关联性很小。

(4) 各个评测指标应该是客观的,并且最好是可以用量化的数据表示的。

5 金融数据中心 SDN 评测指标示例

按照上述的指标选取原则和标准,共选取出 360 多个评测指标。表 2 列出了 SDN 控制器的部分评测指标。

表 2 金融 SDN 控制器部分评测指标

所属维度	一级指标
功能性	支持生成物理拓扑
	支持转发策略制定
	支持南向接口
性能	控制策略下发速率
	Packet-in 处理速度
	拓扑发现时间
	拓扑变化收敛时间
可靠性	支持高可用
	初期故障率
	偶然故障率
可维护性	支持交换机管理
	支持虚拟机管理
可扩展性	交换机对接支持
	物理机对接支持
	支持与云平台对接
安全性	访问控制
	安全审计

6 基于权重的金融数据中心 SDN 评测结果计算

有了评测指标后,已经可以设计测试案例、执行测试,得到相应的测试结果后,即可对测试结果进行评分。但这样仍旧无法通过这些松散的评分对被测对象的整体能力进行直观的评价。因此还需要将分散的结果进行统筹、计算,才能得出对被测对象的整体评价。

在本文中采用加权评分的方式对最终评分进行计算,如式 (1) 所示:

$$V = \sum_{i=1}^n W_i \times T_i \quad (1)$$

式中, V 为评测对象评测值; W_i 为评测维度评测值,计算方法如式 (2) 所示; T_i 为对相应维度的权重。

$$W_i = \sum_{j=1}^m X_j \times T_j \quad (2)$$

式中, X_j 为第 j 个一级评测指标的评测值; T_j 为对相应评测指标的权重。 X_j 的计算方法如式 (3) 所示:

$$X_j = \sum_{k=1}^K X_{jk} \times T_{jk} \quad (3)$$

式中, X_{jk} 为指标 X_j 的第 k 个子评测指标的评测值; T_{jk} 为对相应子评测指标的权重^[11]。

在结果计算中,各个指标的评分值必须在同一分制下,如都在百分制或者十分制下对指标评测结果打分。

7 后续工作规划

如上所述,我们联合产业各方一起,建立了金融数据中心 SDN 评测模型,选定了评测指标以及评分方法。至此,金融数据中心 SDN 基本的评测标准框架已经完成。后续我们会继续开展深度合作,对各项评测指标的权重进行精细化确认,给出各指标的评测权重参考,进一步完善评测标准体系。

8 结语

评测工作是一个产业性、持续性的工作,而评测标准的制定更是一个反复优化的过程。希望未来会有更多的行业同仁能与我们一起,努力将该评测标准做成一套切实可行的金融 SDN 评测方案,共同推动 SDN 在金融行业的落地。

参考文献

- [1] 郑毅, 华一强, 何晓峰, 等. SDN 的特征、发展现状及趋势 [J]. 电信科学, 2013, 29(9): 102.
- [2] 张朝昆, 崔勇, 唐嵩祯, 等. 软件定义网络 (SDN) 研究进展 [J]. 软件学报, 2015, 26(1): 62.
- [3] 田长星. SDN 理论及其在金融业应用前景分析 [J]. 金融科技时代, 2014(7): 65.
- [4] 王酉晟. 浅谈新一代数据中心 SDN 网络技术应用 [J]. 中国金融电脑, 2016(11): 87.
- [5] 宛考, 罗雪峰, 江勇, 等. 软件定义网络系统中面向流的调度算法 [J]. 计算机学报, 2016, 39(6): 1208.
- [6] 刘静, 郭景元, 赖英旭, 等. 基于 OpenFlow 的流量监控架构实践方案 [J]. 北京工业大学学报, 2016, 42(5): 713.
- [7] 赵韶. SDN 在传统银行业数据中心的应用探讨 [J]. 金融科技时代, 2016(8): 24.
- [8] 钟祝君. 招商银行 SDN 网络实践 [J]. 中国金融电脑, 2016(8): 36.
- [9] 顾戎, 王瑞雪, 李晨, 等. 云数据中心 SDN/NFV 组网方案、测试及问题分析 [J]. 电信科学, 2016, 32(1): 126.
- [10] 施海滨, 钟祝君. 基于 SDN 和 NFV 的云安全体系建设 [J]. 中国金融电脑, 2015(11): 26.
- [11] 聂南, 邓璐娟, 夏启明, 等. 面向不同行业软件的质量评测模型及实践 [J]. 计算机科学, 2011, 38(2): 156.