基于深度卷积神经网络的网络流量分类方法

王勇[1,2,3]，周慧怡[2,3]，俸皓[1]，叶苗[3,4]，柯文龙[2]

（1. 桂林电子科技大学计算机与信息安全学院，广西 桂林 541004；2. 桂林电子科技大学信息与通信学院，广西 桂林 541004；
3. 桂林电子科技大学认知无线电与信息处理省部共建教育部重点实验室，广西 桂林 541004；
4.桂林理工大学信息科学与工程学院，广西 桂林 541004）

摘 要：针对传统基于机器学习的流量分类方法中特征选取环节的好坏会直接影响结果精度的问题，提出一种基于卷积神经网络的流量分类算法。首先，通过对数据进行归一化处理后映射成灰度图片作为卷积神经网络的输入数据，然后，基于 LeNet-5 深度卷积神经网络设计适于流量分类应用的卷积层特征面及全连接层的参数，构造能够实现流量的自主特征学习的最优分类模型，从而实现网络流量的分类。所提方法可以在避免复杂显式特征提取的同时达到提高分类精度的效果。通过公开数据集和实际数据集的系列仿真实验测试结果表明，与传统分类方法相比所提算法基于改进的 CNN 流量分类方法不仅提高了流量分类的精度，而且减少了分类所用的时间。

关键词：流量分类；卷积神经网络；归一化；特征选择

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018018

Network traffic classification method basing on CNN

WANG Yong[1,2,3], ZHOU Huiyi[2,3], FENG Hao[1] , YE Miao[3,4], KE Wenlong[2]

1. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
2. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China
3. Key Laboratory of Cognitive Radio and Information Processing, Guilin University of Electronic Technology, Guilin 541004, China
4. Information Science and Technology, Guilin University of Industrial Technology, Guilin 541004, China

Abstract: Since the feature selection process will directly affect the accuracy of the traffic classification based on the traditional machine learning method, a traffic classification algorithm based on convolution neural network was tailored. First, the min-max normalization method was utilized to process the traffic data and map them into gray images, which would be used as the input data of convolution neural network to realize the independent feature learning. Then, an improved structure of the classical convolution neural network was proposed, and the parameters of the feature map and the full connection layer were designed to select the optimal classification model to realize the traffic classification. The tailored method can improve the classification accuracy without the complex operation of the network traffic. A series of simulation test results with the public data sets and real data sets show that compared with the traditional classification methods, the tailored convolution neural network traffic classification method can improve the accuracy and reduce the time of classification.

Key words: network traffic classification, convolutional neural network, normalized, feature selection

收稿日期：2017-10-12；**修回日期**：2017-12-19

通信作者：俸皓，fengh@guet.edu.cn

基金项目：国家自然科学基金资助项目（No.61662018, No.61661015）；中国博士后科学基金资助项目（No.2016M602922XB）；广西自然科学基金资助项目（No.2016GXNSFAA380153）；桂林电子科技大学研究生教育创新计划基金资助项目（No.2018YJCX53, No.2018YJCX20）；桂林理工大学科研启动基金资助项目（No.GUTQDJJ20172000019）

Foundation Items: The National Natural Science Foundation of China (No.61662018, No.61661015), Project Funded by China Postdoctoral Foundation (No.2016M602922XB), The Natural Science Foundation of Guangxi Autonomous Region (No.2016GXNSFAA380153), Innovation Project of Guest Graduate Education (No.2018YJCX53, No.2018YJCX20), Foundation of Guilin University of Technology (No.GUTQDJJ20172000019)

1　引言

随着大数据时代的到来，新型网络应用不断兴起，网络的组成越来越复杂。如何从日益增长的流量数据中获得更有价值的信息已经成为各大运营商制订发展规划不可或缺的评估指标之一。网络流量分类作为网络管理与网络安全的关键技术之一，不但能够优化网络配置，降低网络安全隐患，而且能够根据用户的行为分析提供更好的服务质量。目前，基于机器学习的流量分类方法一直是研究者关注的热点，应用在网络流量分类领域的机器学习算法可以分为浅层学习和深度学习[1]2 种。

浅层学习主要包括：支持向量机、决策树[2,3]、贝叶斯[4]和 *k*-means[5]等。文献[6]创建了一个跨越多天的 WLAN 数据集，包括多种流量类型和应用程序，从数据中提取 63 个特征并用于训练 6 种不同的机器学习算法。文献[7]提出了使用流程级特征实现在线流量分类的算法和架构。首先，设计了基于 C4.5 决策树算法和熵 MDL（minimum description length）离散化算法的流量分类器对 8 个主要应用进行分类，总体精度达到 97.92%。其次，将离散化算法和 FPGA（现场可编程门阵列）以及多核平台分类器进行合并的方法，实现分类器的优化。无论文献[6]还是文献[7]均是通过人工选择和组合特征的方法实现最终的流量分类，增加了网络流量分类的工作量。文献[8]使用小波领导者多分形式主义（WLMF, wavelet leaders multifractal formalism）从网络流中提取多重分形特征来描述网络流量，然后，将基于主成分分析（PCA, principal component analysis）的特征选择方法应用于这些多重分形特征以消除不相关和冗余特征。实验结果表明与现有的基于机器学习的方法中研究的传输层数据特征相比，支持向量机（SVM, support vector machine）的分类准确性显著提高。文献[9]提出了一种 SPP-SVM 的实时精确 SVM 训练模型。该模型首先通过 PCA 对数据进行特征提取，降低原有特征的维度，然后，通过采用一种改进的粒子群优化算法自动搜索核函数的最优工作参数，使其与传统的 SVM 相比，可以通过少量的训练样本就能显著的提高流量分类的精度。文献[9]还通过同时对所有数据进行数值缩放操作来减小工作计算量。文献[8,9]均采用了 PCA 降维的方法对网络流量数据的特征进行降维后再实现网络流量分类，本文也采用了 PCA 算法对 Moore 数据集进行特征降维操作后与本文算法进行了对比。

近年来，深度学习在图像识别[10]、语音识别[11]、音频处理[12]和自然语音处理[13]等各大领域均有应用并取得了很好的成绩。深度学习主要包括深度置信网络、卷积神经网络和递归神经网络等。文献[14]通过使用 WK-ELM 算法中的 GA 技术来获取神经网络最合适的参数值，以及使用极端学习机的优化方法所提供隐含层的单元数量，来增加正确分类的准确率。该方法很好地解决了神经网络中参数选取的问题，针对 Moore 数据集进行流量分类达到了 96.57%的准确率。但是该分类方法是一个静态的分类过程，需要根据当前样本人工选取节点数和参数，即只能一次性地完成训练，一旦训练过程样本发生变化，就要重新开始训练，无法动态更新参数和训练样本。面对动态的批量的分类过程，会浪费大量的时间来更新参数和样本。文献[15]针对 P2P 流量分类准确率较低的问题,提出一种基于深度学习结构、半监督的深度置信网络（DBN, deep belief networks）流量分类方法，构造 P2P 流量合适的特征空间，建立基于 DBN 的网络流量分类模型，并对模型的隐含节点个数和隐含层个数进行选择，提高 DBN 模型对 P2P 流量的分类准确率。但是该方法在构造数据集的阶段就需要进行特征提取，且采用 DBN 模型进行流量分类计算过于复杂，难以满足实际应用中的实时性需求。文献[14,15]在实验中取得的成功，证明了深度学习在流量分类上的可行性。相对于经典的机器学习方法，在训练数据发生变化的时候，基于深度学习的分类模型可以在其原有最优模型的基础上进行微调，生成新的最优分类模型完成流量分类。而基于机器学习的分类模型需要对变化的训练数据重新进行特征提取和训练学习，完成对新数据的流量分类。

针对以上研究中出现的问题，本文提出了基于离差标准化的卷积神经网络（MMN-CNN, min-max normalization convolutional neural network），该方法可以隐性地从训练数据中进行网络学习并提取特征，不仅避免了人工特征选取的麻烦，很好地解决了不同分类算法对特征选取的差异性，而且提高了流量分类的精度。

2　流量分类问题及描述

网络流量分类的总体流程包括网络数据的采集，带有准确背景信息数据集的生成，数据集的预处理，流量特征的提取以及分类。为了验证所提算法的可行性，本文分别采用现有的数据集和通过网络协议数据分析工具（Microsoft Network Monitor）采集的流量数据集进行实验。目前，针对数据集进行分类的方法，常用的特征选择方法的基本流程如图 1 所示，根据不同的搜索策略对原始特征集生成特定的特征子集后依据某种约定进行特征子集评估，最终得到最优的特征子集。这种方法不仅需要额外的计算开销而且需要根据分类应用的特点来选取合适的特征。针对上述方法存在的问题，本文采用卷积神经网络对原始数据集进行特征的自主学习，通过多隐含层的深层结构来构造特征空间，对大量数据的自主学习发现数据的特征表示来构造合适的特征空间。该方法不仅解决了特征子集选取的困难而且提高了分类的效率，为网络流量进行实时分类奠定了基础。

卷积神经网络的输入形式一般为二维矩阵，如何设计较优的数据集预处理方法将影响整个分类的效果。本文选取了 2 种不同数据集作为输入数据，第一种为 Moore 数据集[16]，一共包括 12 个流量类别，每条流量中包含 249 流量特征，其中，最后一个特征是每条流量相对应的类别，应用类型为 WWW 的具体数据格式如图 2 所示。

第二种为实际数据集，一共包括 4 类，每条流量包含 784 位流量数据，其中每条流量的最后一位是对应的应用类别，应用类型为通信类的具体数据格式如图 3 所示。

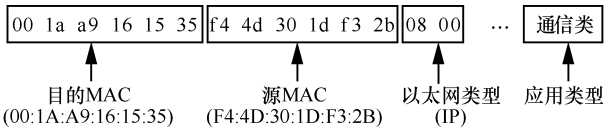


图 3　应用类型为通信类的数据格式

首先，对数据集进行数据预处理将网络流量数据转化为灰度图片，然后，将灰度图片作为卷积神经网络的输入数据进行学习。基于 MMN-CNN 的分类方法是将数据集中所有的数据进行归一化处理，然后，根据数据集中每条数据的长度构造合适的二维矩阵维度，将每一位数据对应为二维矩阵中的一个元素，进而完成灰度图片映射。

3　基于改进卷积神经网络的网络流量分类方法

采用 MMN-CNN 对网络流量进行分类的核心思想是通过对数据集进行预处理，再利用卷积神经网络对预处理后的数据进行自主特征学习，从而完成对流量的分类。其整体架构如图 4 所示。

3.1　网络流量数据集构建

带有标签数据集是保证深度学习性能的一个关键因素，为此本节先对需要处理的原始网络流量数据及标签的标注工作做简单介绍，为了验证后面所设计的基于深度学习的流量分类方法的性能，使用了公开数据集和实际流量数据集，其采集和标注工作如下。

1) Moore 数据集

Moore 数据集是由 1 000 个用户通过一条千兆全双工以太网链路连接到互联网，使用网络监测器

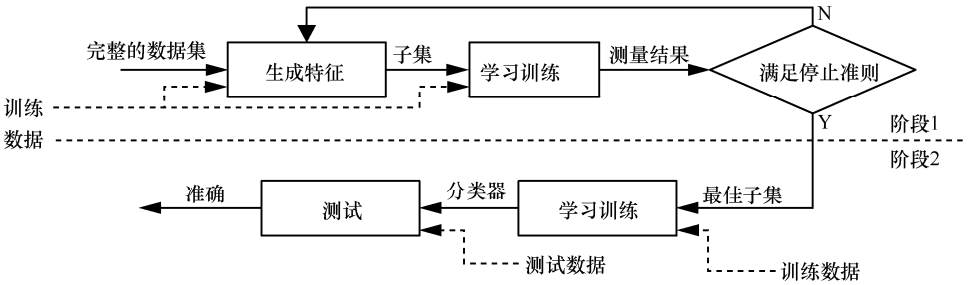


图 1　特征选择的基本流程

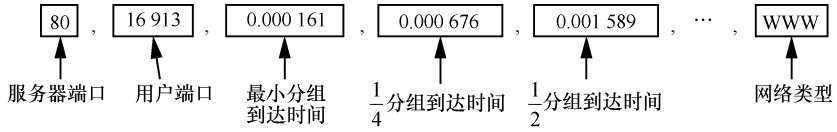


图 2　应用类型为 WWW 的数据格式

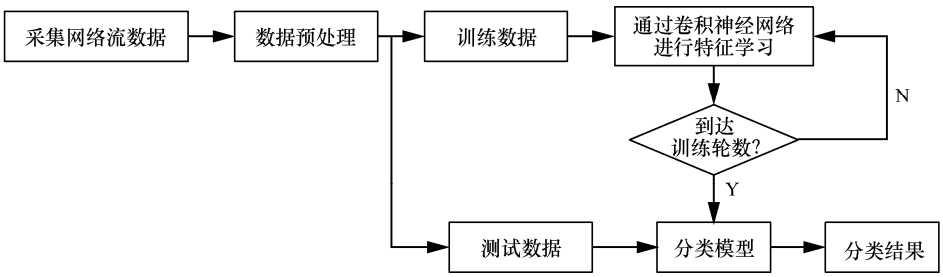


图 4　网络流量分类整体架构

采集 24 h 的网络流量，然后采用抽样算法在 24 h 的流量中得到 377 526 个网络样本，并根据应用类型分为 12 种类型。Moore 数据集中每条网络流样本都是从一条完整的 TCP 双向流抽象出来，包含 249 项属性，其中，最后一项属性是每条网络流相对应的类别。Moore 流量数据集统计信息如表 1 所示。

表 1　　Moore 流量数据统计信息

类别	数量	比例
WWW	328 092	86.906%
MAIL	28 567	7.567%
FTP-DATA	5 797	1.536%
FTP-PASV	2 688	0.712%
FTP-CONTROL	3 054	0.809%
SERVICES	2 099	0.556%
DATABASE	2 648	0.701%
P2P	2 094	0.555%
ATTACK	1 793	0.475%
MUITIMEDIA	576	0.153%
INTERACTIVE	110	0.028%
GAME	8	0.002%
总计	377 526	100.000%

2) 实际数据集

本文所采用的实际数据集是通过微软发布的一款网络协议数据分析工具 Microsoft Network Monitor 对校园网流量常用的 10 种应用软件进行数据采集得到 679 207 个网络样本，并根据 10 种应用软件的相关性分为 4 种类型，分别为网页类（Chrome、FireFox 和 360）、通信类（QQ）、下载类（Thunder 和 BaiduNetDisk）以及视频类（IQIYI、mgtv、YOUKU 和 QQLive）。其中，一条具体应用类型为 BaiduNetDisk 的数据采集结果如图 5 所示，实验中构造的数据集主要由应用标签和图 5 中的十六进制的流量数据组成。首先，将流量数据进行固定长度的截取，本文将流量数据分为 784 个字段，每一个字段由 8 bit 构成，转化为十进制即为 0~255 之间的数值。然后，将每条数据流最后一个字段加上对应的应用标签，即采用第一栏捕获的流量的应用类型。

```
0000 00 1A A9 16 15 35 40 8D 5C 5F 36 AD 08 00 45 00
0010 00 63 06 53 40 00 80 06 00 00 CA C1 35 0D B4 95
0020 91 F1 26 B8 01 BB 3C 10 AE 6D B5 B9 2C 2F 50 18
0030 40 E9 46 AB 00 00 14 03 01 00 01 01 16 03 01 00
0040 30 8D 82 D3 4F 68 77 2F 0E 3B 83 2E 10 B4 7A FC
0050 CC 34 60 CF 13 E4 AD E8 D3 A7 BA 16 F3 2B 42 7F
0060 62 0A AE E7 42 CC DD F4 38 50 7D 80 DC C0 AC 66
0070 C7
```

图 5　一条应用类型为 BaiduNetDisk 的采集数据

由 Microsoft Network Monitor 采集的实际网络流量数据统计信息如表 2 所示。

表 2　　实际流量数据统计信息

应用类型	数量	比例
Chrome	69 851	10.28%
FireFox	59 712	8.79%
360	62 726	9.24%
QQ	91 600	13.49%
Thunder	62 207	9.16%
BaiduNetDisk	58 626	8.63%
IQIYI	90 547	13.33%
mgtv	66 368	9.77%
YOUKU	59 674	8.79%
QQLive	57 896	8.52%
总计	679 207	100.00%

本文首先将数据集分成训练数据集和测试数据集 2 个部分，这 2 个数据集中每一种类别的比例与原流量数据保持一致，随机选取 100 000 条数据作为测试数据集，其他为训练数据集。

3.2　数据预处理

本文根据卷积神经网络能够隐性地训练数据中进行网络学习的特点，对数据集进行预处理的过

程设计如下。

为了消除网络流量之间的量纲关系，使数据之间具有可比性，本文首先对数据集中的每一条相对应的特征进行数值归一化。具体过程如下。

假设数据集可以表示为 n 行 m 列矩阵

$$\boldsymbol{T}=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix} \tag{1}$$

那么根据特征来划分可以将该矩阵表示为

$$\boldsymbol{B}_i=\left[A_{1i},A_{2i},A_{3i},\cdots A_{ni}\right]^{\mathrm{T}} \tag{2}$$

其中，$\boldsymbol{B}_i$ 表示第 i 个特征在所有数据中的值，那么矩阵 $\boldsymbol{T}$ 可以表示为

$$\boldsymbol{T}=\left[\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3,\cdots,\boldsymbol{B}_m\right] \tag{3}$$

对每一列进行如下归一化处理后得到 $\boldsymbol{B}_i'$

$$\boldsymbol{B}_i'=\frac{\boldsymbol{B}_i}{\max\{A_{1i},A_{2i},A_{3i},\cdots,A_{ni}\}} \tag{4}$$

那么量化后的矩阵可以表示为

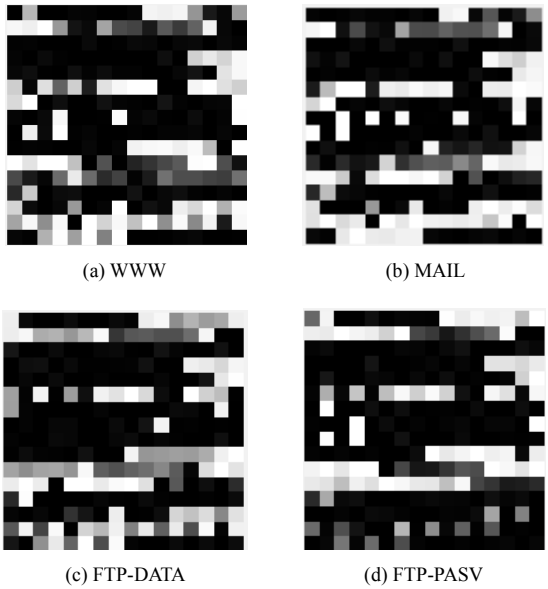
$$\boldsymbol{T}'=\left[\boldsymbol{B}_1',\boldsymbol{B}_2',\boldsymbol{B}_3',\cdots,\boldsymbol{B}_m'\right] \tag{5}$$

以 Moore 数据集为例，介绍以上设计的数据预处理过程。根据数据集中的特征数量，构建适合卷积神经网络学习的矩阵维度。根据 Moore 数据集具有的 249 位特征，构建一个 16×16 的矩阵，由于特征维数少于矩阵元素个数，故在矩阵的末位进行相应的补 0 操作。

将构建好的矩阵中的每个元素作为一个像素点，矩阵中的值作为像素的灰度。元素值越大灰度越大，图片越接近白色；反之，则灰度越小越接近黑色。每一条网络流量均将转换成对应的一张灰度图片，图 6 为 Moore 数据集中比例较高的 4 类应用类型所对应的灰度图片。图 7 为实际数据集中应用类型所对应的灰度图片。

3.3 卷积神经网络结构

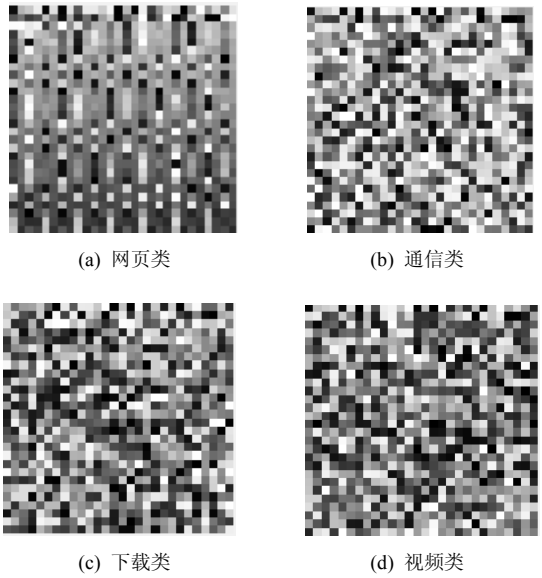
卷积神经网络[17]的基本结构由输入层、卷积层、池化层、全连接层、输出层构成。一般会设计成若干个卷积层和池化层交替模型，即一个卷积层连接一个池化层，池化层后再连接一个卷积层，依此类推。全连接层常用于连接最后 2 层节点，用于输出最终结果。



(a) WWW (b) MAIL

(c) FTP-DATA (d) FTP-PASV

图 6 不同 Moore 流量数据对应的灰度图片



(a) 网页类 (b) 通信类

(c) 下载类 (d) 视频类

图 7 实际数据对应的灰度图片

1) 卷积层

卷积层是卷积神经网络中最重要的一部分，又被称为过滤器或内核，由多个特征面组成，每个特征面由多个神经元组成。卷积层中每一个节点的输入只是上一层神经网络的一个小块，这个小块的长和宽都是人为指定，叫作卷积核，常用的卷积核大小有 3×3 或 5×5。卷积层试图将神经网络中的每一小块进行更加深入地分析从而得到抽象程度更高的特征。卷积层的形式为

$$x_j^l=f\Big(\sum_{i\in M_j}x_j^{l-1}k_{ij}^l+b_j^l\Big) \tag{6}$$

其中，l 为当前层，b 为当前层的偏置，k 为卷积核，M_j 为第 j 个卷积核对应的卷积窗口，激活函数通常采用 sigmoid、tanh 或 ReLU 等函数。

2) 池化层

池化层也称为采样层，紧跟在卷积层之后，同样由多个特征面组成，它的每一个特征面唯一对应于上一层的一个特征面。池化层神经网络不会改变上一层特征面的个数，但是它可以缩小每一个特征面的大小。通过池化层，可以进一步缩小最后全连接层中节点的个数，从而达到减少整个神经网络中参数的目的。池化层的形式为

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \tag{7}$$

其中，$down(\cdot)$ 为次抽样函数。次抽样函数通常是对输入图像中每一个 $n \times n$ 大小的区域加权求和，故输出图像的大小变为输入图像大小的 $\frac{1}{n}$，输出的特征映射图有自己的网络乘性参数 β 和偏置 b。

3) 全连接层

全连接层通常位于卷积神经网络模型的最后位置，作用是计算网络的最终输出结果。分类任务中通常在这一层中训练一个分类器，将学习到的高层特征作为分类器的输入，输出结果为分类的结果。

3.4 MMN-CNN 的训练过程

MMN-CNN 模型的训练过程主要涉及网络的前向传播和反向传播，前向传播用于得到预测值，体现了特征信息的传递，而反向传播则是用于更新变量，体现了误差信息对模型参数的矫正。

1) 前向传播

卷积神经网络的前向传播与普通神经网络的前向传播过程一样。卷积层的前向传播形式为

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} k_{ij}^l + b_j^l) \tag{8}$$

其中，本文中使用激活函数为 ReLU 函数。

2) 反向传播

卷积神经网络的反向传播算法和 BP 神经网络类似。首先，本文使用的总体代价函数定义为

$$E^n = -\frac{1}{n}\sum_{xj}[y_j \ln a_j^l + (1-y_j)\ln(1-a_j^l)] + \frac{\lambda}{2n}\sum_{\omega}\omega^2 \tag{9}$$

其中，右边第 1 项为常规的交叉熵的表达式，第 2 项为正则化项，使用一个正则化参数对权重的平方和进行量化调整，本文使用的正规化参数为 0.01。反向传播算法是根据定义好的损失函数优化卷积神经网络中的参数的取值，从而使卷积神经网络模型在训练数据集上的损失函数达到一个较小值。

权值参数的调整方向为

$$\omega \to \omega - \eta \frac{\partial E^n}{\partial \omega} \tag{10}$$

偏置参数的调整方向为

$$b \to b - \eta \frac{\partial E^n}{\partial b} \tag{11}$$

其中，η 为学习速率，本文中使用的学习速率为 0.55。

3.5 改进的卷积神经网络结构

随着人工智能和大数据时代的到来以及计算能力的飞速发展，深度学习已经成为各界研究的热点。作为深度学习的一种重要模型——卷积神经网络，最初是为解决图像识别等问题而设计并取得了很好的效果，但在网络流量分类上却少有人涉及。因为无论是图像还是流量数据均是由数值构成，故本文选取卷积神经网络作为特征学习及分类的方法，用于对网络流量进行分类。

LeNet-5 模型是文献[18]在 1998 年设计的用于手写数字识别的卷积神经网络，是早期卷积神经网络中最有代表性的实验系统之一。LeNet-5 模型总共有 7 层，包括输入层、卷积层、池化层、卷积层、池化层、全连接层和输出层。针对网络流量数据的特点，本文对传统的 LeNet-5 模型进行了以下改进。

1) 根据对数据集的分析得到合适大小的灰度图片，将网络的输入层设计为 16×16 的矩阵。

2) 网络流量分类的目的是将数据集中不同的流量应用类型进行分类，其中，包括 12 种不同的应用类型。因此，将传统的 LeNet-5 模型的输出层由 10 个神经元改为 12 个神经元。

3) 根据网络流量数据的特点，设计了 6 种具有不同结构的卷积神经网络。具体所设计的 6 种不同网络结构的卷积神经网络如表 3 所示。

针对设计过程所遇到的问题有以下说明。1) 由于输入层的数据样本大小为 16×16，计算量难度不大，为避免图像的边缘信息丢失得太快，在当前层矩阵的边界加入全 0 填充，使卷积前后的图像尺寸保持相同，可以保持边界的信息，且本文中卷积核的移动步长为 1。2) 对于一个 5 层的卷积神经网络而言，当卷积核的特征面大于 64 时，容易出现局

表 3　　6 种不同网络结构的卷积神经网络

编号	C_1 卷积层		S_2 池化层		C_3 卷积层		S_4 池化层		C_5 全连接层	
	卷积核	输出	采样窗口	输出	卷积核	输出	采样窗口	输出	卷积核	输出
1	32×(3×3)	32×(16×16)	2×2	32×(8×8)	64×(5×5)	64×(8×8)	2×2	64×(4×4)	256×(4×4)	256×1
2	32×(3×3)	32×(16×16)	2×2	32×(8×8)	64×(5×5)	64×(8×8)	2×2	64×(4×4)	128×(4×4)	128×1
3	16×(3×3)	16×(16×16)	2×2	16×(8×8)	32×(5×5)	32×(8×8)	2×2	32×(4×4)	256×(4×4)	256×1
4	16×(3×3)	16×(16×16)	2×2	16×(8×8)	32×(5×5)	32×(8×8)	2×2	32×(4×4)	128×(4×4)	128×1
5	8×(3×3)	8×(16×16)	2×2	8×(8×8)	16×(5×5)	16×(8×8)	2×2	16×(4×4)	256×(4×4)	256×1
6	8×(3×3)	8×(16×16)	2×2	8×(8×8)	16×(5×5)	16×(8×8)	2×2	16×(4×4)	128×(4×4)	128×1

部最优的情况，故所设计的特征面均不大于 64。

将所设计的 6 种不同卷积层结构的卷积神经网络分别采用 2 种不同池化层进行性能对比，常用的池化层主要有 2 种，一种是最大池化，另一种是平均池化。其实验检测结果如表 4 和表 5 所示。基于 MMN-CNN 流量分类方法，分为训练和测试 2 个阶段，对于后期开展实时流量分类工作，由于训练阶段可以进行离线分析，并不影响最后的分类效果，因此，本文只针对各种算法的测试时间进行对比分析。

表 4　　采用最大池化方法时的检测结果

编号	总体准确率	测试时间/s
1	86.90%	20.13
2	98.90%	18.41
3	99.18%	12.01
4	98.98%	11.30
5	99.32%	6.29
6	99.30%	6.03

表 5　　采用平均池化时的检测结果

编号	总体准确率	测试时间/s
1	98.30%	20.69
2	98.33%	18.67
3	98.42%	11.55
4	98.40%	10.77
5	98.22%	6.22
6	98.58%	5.99

实验结果表明，总体而言，逐层增加特征面的数量，分类准确率变化不大，而测试时间增大的特别明显。在采用最大池化方法时，编号为 1 的模型陷入了局部最优，而随着特征面的减少，分类准确率得到了很好的提升。对比以上 2 种检测结果，编号为 6 的模型不仅在分类准确率上而且在测试时间上均取得了很好的效果，故本文选取采用最大池化方法编号为 6 的模型为最优的卷积神经网络。该卷积神经网络模型的总体准确率和测试时间分别为 99.30%和 6.03 s。

根据上述实验结果选取的编号为 6 的最优 MMN-CNN 模型的网络结构如图 8 所示。C_1 层为卷积层，采用全 0 填充后使用感知阶段预训练得到的 8 个 3×3 大小的卷积核对输入数据进行卷积，得到 8 张 16×16 的特征面。S_2 层为池化层，使用 2×2 大小的窗口对 C_1 层的 8 张特征面池化得到 8 张 8×8 的特征面。C_3 层为卷积层，采用全 0 填充后使用预训练得到的 16 个 5×5 大小的卷积核对输入数据进行卷积，得到 16 张 8×8 的特征面。S_4 层为池化层，使用 2×2 大小的窗口对 C_3 层的 16 张特征面池化得到 16 张 4×4 的特征面。F_5 层为全连接层，利用 128 个 4×4 大小的卷积核将 16 张 4×4 的特征面映射为一个 128×1 的向量。输出层利用 soft-max 分类器，输出结果为 12 类。

4　实验测试与结果分析

为了验证 MMN-CNN 算法的可行性，本文实验环境参数如表 6 所示。

表 6　　实验环境参数

类别	参数
硬件环境	Supermicro X9DRG-QF
操作系统	Windows 7,64 bit
处理器	Intel Xeon E5-2630 2.30 GHz
内存	64 GB DDR3 1 333 MHz
Anaconda3	4.1.1 版本 64 bit
Tensorflow	1.1.0 版本

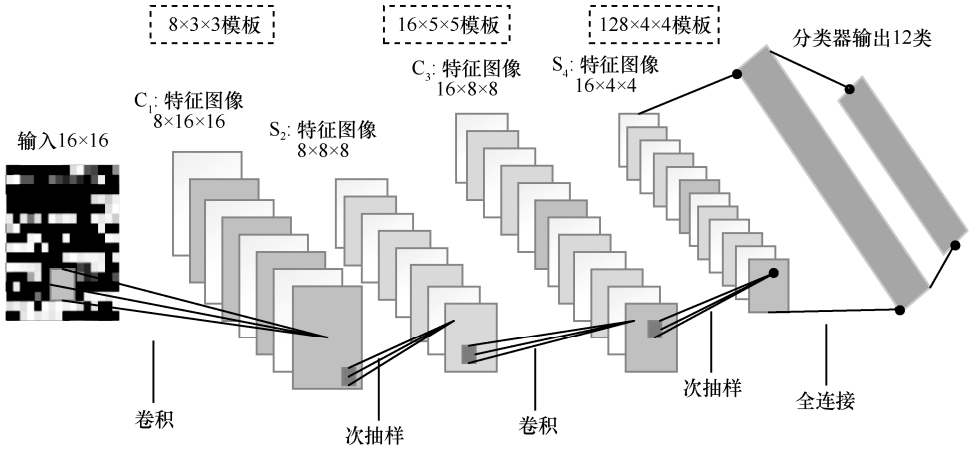


图 8　本文设计的卷积神经网络结构

实验结果的评价指标主要分为 2 个部分：1) 分类的准确率，包括每一类应用类型的准确率，可信度和所有类别的整体准确率[19]；2) 检测训练好的卷积神经网络模型的测试时间。具体而言，在本文实验中，类 i 的准确率 $A_i=\dfrac{TP_i}{TP_i+FN_i}$，类 i 的可信度 $T_i=\dfrac{TP_i}{TP_i+FP_i}$，整体准确率 $OA=\dfrac{\sum_{i=1}^{m}TP_i}{\sum_{i=1}^{m}(TP_i+FH_i)}$，其中，$TP_i$ (true positive)指的是实际类型为 i 的样本中被分类模型正确预测的样本数。FN_i (false negative)指的是实际类型为 i 的样本文被分类模型误判为其他类型的样本数。FP_i (false positive)指的是实际类型为非 i 的样本中被分类模型误判为类型 i 的样本数量。

常用降维方法有主成分分析（PCA）、高斯随机映射（GRP, Gaussian random projection）和稀疏随机映射（SRP, sparse random projection）等。相比于高斯随机映射，稀疏随机映射更能保证降维的质量，并提高内存的使用效率和运算效率。为了验证所提出的 MMN-CNN 算法在流量分类问题上的优势，因此，针对 Moore 数据集，采用了 PCA 和 SRP 对流量数据降维至 16 个特征。接着使用文献[20]采用的数据预处理方法将选取网络流量数据的特征并生成特征向量，然后将流特征向量标准化，进而利用特征之间的欧拉距离将流特征向量转换为欧拉矩阵并可视化为灰度图片。本文将针对这 2 种不同灰度图片的映射方法，使用 3.4 节中选取的最优卷积神经网络模型，分别根据评价指标的各个参数进行实验对比，具体实验结果如表 7~表 9 所示。

表 7　采用不同算法的总体准确率和测试时间

方法	总体准确率	测试时间/s
PCA	96.09%	6.03
SRP	97.23%	5.92
本文算法	99.30%	6.03

表 8　采用不同算法的类准确率

方法	PCA	SRP	本文算法
WWW	99.57%	99.74%	99.83%
MAIL	91.13%	94.79%	99.39%
F-D	73.15%	93.04%	99.8%
F-P	2.81%	48.53%	96.63%
F-C	75.34%	75.59%	95.92%
SERV	91.20%	96.05%	98.38%
DB	0%	0.43%	82.48%
P2P	17.45%	23.92%	90.45%
ATT	65.13%	67.23%	66.32%
MULT	0%	0%	54.25%
INT	0%	0%	0%
GAME	0%	0%	0%

表 9　采用不同算法的类可信度

方法	PCA	SRP	本文算法
WWW	96.97%	98.07%	99.61%
MAIL	90.56%	92.91%	98.61%
F-D	93.83%	94.33%	98.90%
F-P	29.41%	55.18%	92.24%
F-C	73.79%	92.18%	96.16%
SERV	90.07%	96.75%	96.81%
DB	0%	23.08%	98.81%
P2P	71.85%	70.00%	83.11%
ATT	92.81%	94.96%	99.06%
MULT	0%	0%	80.58%
INT	0%	0%	0%
GAME	0%	0%	0%

从上述实验结果可以得出，本文所提MMN-CNN 算法，直接对 Moore 数据集进行数据预处理后转换成灰度图片，相较于采用 PCA 和 SRP 降维方法，无论在总体准确率还是针对不同应用类型的准确率和可信度进行比较，均有大幅度提升。且相较于文献[14]对 Moore 数据集进行流量分类的准确率 96.57%，本文提出的 MMN-CNN 算法的总体准确率提高了 2.73%。因此，卷积神经网络自主进行特征选取不仅减少了工作量而且提高了分类的准确率。从类准确率和类可信度的实验结果发现， INTERACTIVE 和 GAME 这 2 种类型的准确率均为 0，说明了卷积神经网络对数据量少的应用类型分类效果并不理想。但是，随着网络技术的日新月异，网络流量呈 EB 级增长，所以针对实际网络流量分析，这个问题能得到很好解决。

为了进一步验证 MMN-CNN 自主特征学习的优越性，本文采用微软发布的 Microsoft Network Monitor 对校园网流量进行应用层数据抓取，并根据相应的类型打上应用标签。由于卷积神经网络的数据输入格式大小是固定的，故对所抓取的流量数据进行固定大小的截断，其次输入数据中每个元素有由 0~255 之间的数值组成。本文针对实际流量数据的特点将每一条数据流量构造为 28×28 的二维矩阵，然后采用 MMN-CNN 对其进行自主特征学习完成流量的分类，具体实验结果如表 10 所示。

表 10　　实际流量数据的分类结果

应用类型	类准确率	类可信度
网页	92.89%	80.07%
通信	85.65%	95.37%
下载	99.36%	99.84%
视频	86.99%	94.10%

通过实验可以看出，使用 MMN-CNN 对未经过特征提取的流量数据进行学习，总体正确率可以达到 90.68%，并且每一种应用类型的分类准确率均取得了较好的效果。MMN-CNN 方法在流量分类上取得的成功为接下来实时流量分类奠定了基础。

5 结束语

本文首先介绍了常用的网络流量分类方法，然后在基于相关研究的基础上，给出了一种流量数据预处理的 MMN-CNN 算法，应用于对网络流量数据的特征学习过程，提取出更加抽象的特征，然后通过设计不同的特征面及全连接层的参数完成最优分类模型的选取，从而实现流量分类。为了验证该算法的可行性，分别通过对现有的流量数据集和抓包软件对所采集的数据集进行实验分析，将原始数据直接输入到 MMN-CNN 网络中，避免了人工特征提取导致的误差积累，提高了流量分类的效果。实际网络流量数据愈发的呈现多源异构，深度学习应用于网络流量分类问题的研究还处于起步阶段。在将来的工作中，本文将针对以下 3 个方面进行更进一步的研究：1) 利用深度学习的其他算法解决网络流量数据大小不一致的分类方法研究；2) 利用大数据分析平台实现算法的分布式计算，提高流量分类的效率；3) 增加卷积神经网络的层数，比较实际网络流量分类的效果。

参考文献：

[1] LENCUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 512:436-444.

[2] DING J, HUANG L, YUPNEG T, et al. A flow nature classification method based on multi-features of *n*-gram[J]. Computer Applications & Software, 2017.

[3] GUANG L, RONGHUA G. Cascaded classifier for improving traffic classification accuracy[J]. IET Communications,2017,11: 1751-1758.

[4] SCHMIDT B, AL-FUQAHA A, GUPTA A. Optimizing an artificial immune system algorithm in support of flow-based internet traffic classification[J]. Applied Soft Computing, 2017,54: 1-22.

[5] DONG Y N, ZHAO J J, JIM J. Novel feature selection and classification of Internet video traffic based on a hierarchical scheme[J]. Computer networks, 2017,119: 102-111.

[6] KORNCKY J, ABDUL-HAMEED O, KONDOZ A. Radio frequency traffic classification over WLAN[J]. IEEE-ACM Transactions on Networking, 2017,25: 56-68.

[7] TONG D, QU Y R, PRASANNA V K. Accelerating decision tree based traffic classification on FPGA and multicore platforms[J]. IEEE Transactions on Parallel & Distributed Systems, 2017:1.

[8] Shi H, Li H, Zhang D, et al. Efficient and robust feature extraction and selection for traffic classification[J]. Computer Networks the International Journal of Computer & Telecommunications Networking, 2017, 119(C):1-16.

[9] CAO J, FANG Z Y, QU G N, et al. An accurate traffic classification model based on support vector machines[J]. Networks, 2017,27.

[10] HE K M, ZHANG X Y, REN S Q. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C]//IEEE International Conference on Computer Vision. 2015: 1026-1034.

[11] SUN S N, ZHANG B B, XIE L, et al. An unsupervised deep domain adaptation approach for robust speech recognition[J]. Neurocomputing,

2017,257: 79-87.

[12] WILLIAMSON D S, WANG D L. Time-frequency masking in the complex domain for speech dereverberation and denoising[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2017, 25: 1492-1501.

[13] ZHANG Y, MARSHALL I, WALLACE B C. Rationale-augmented convolutional neural networks for text classification[C]//Conf Empir Methods Nat Lang Process. 2016:795.

[14] ERTAM F, AVCI E. A new approach for internet traffic classification: GA-WK-ELM[J]. Measurement, 2017, 95:135-142.

[15] 白雪. 基于 DBN 的网络流量分类的研究[D]. 呼和浩特: 内蒙古大学, 2015.
BAI X. Research on Internet traffic classification using DBN[D]. Hohehot: Inner Mongolia University, 2015.

[16] MOORE A W, ZUEV D. Discriminators for use in flow-based classification[R]. Technical report, Intel Research, Cambridge, 2005.

[17] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1-23.
ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1-23.

[18] LENCUN Y, BOTTOU L, BENGIO Y. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 862: 2278-2324.

[19] 徐鹏, 林森. 基于 C4.5 决策树的流量分类方法[J]. 软件学报, 2009, 20(10):2692-2704.
XU P, LIN S. Internet traffic classification using C4.5 decision tree[J]. Journal of Software, 2009, 20(10):2692-2704.

[20] 寇广, 汤光明, 王硕, 等. 深度学习在僵尸云检测中的应用研究[J]. 通信学报, 2016, 37(11):114-128.
KOU G, TANG G M, WANG S, et al. Using deep learning for detecting BotCloud[J]. Journal on Communications, 2016, 37(11):114-128.

[作者简介]

王勇（1964-），男，四川阆中人，博士，桂林电子科技大学教授、博士生导师，主要研究方向为云计算、网络流量分析、信息安全。

周慧怡（1993-），女，湖南永州人，桂林电子科技大学硕士生，主要研究方向为大数据分析。

俸皓（1978-），男，广西桂林人，博士，桂林电子科技大学副教授，主要研究方向为无线传感器网络、嵌入式系统。

叶苗（1977-），男，广西桂林人，博士，桂林电子科技大学教授、硕士生导师，主要研究方向为网络计算、无线传感器网络、模式识别与图像处理。

柯文龙（1989-），男，安徽铜陵人，桂林电子科技大学博士生，主要研究方向为大数据分析。