基于 Spark 框架和 PSO 优化算法的电力通信网络安全态势预测

金 鑫[1] 李龙威[1] 苏国华[2] 刘晓蕾[2] 季佳男[3]
(中央财经大学信息学院 北京 100081)[1] (北京国电通网络技术有限公司 北京 100070)[2]
(人力资源和社会保障部人事考试中心 北京 100011)[3]

摘 要 随着电力通信网络规模的不断扩大,电力通信网络不间断地产生海量通信数据。同时,对通信网络的攻击手段也在不断进化,给电力通信网络的安全造成极大威胁。针对以上问题,结合 Spark 大数据计算框架和 PSO 优化神经网络算法的优点,提出基于 Spark 内存计算框架的并行 PSO 优化神经网络算法对电力通信网络的安全态势进行预测。本研究首先引入 Spark 计算框架,Spark 框架具有内存计算以及准实时处理的特点,符合电力通信大数据处理的要求。然后提出 PSO 优化算法对神经网络的权值进行修正,以增加神经网络的学习效率和准确性。之后结合 RDD 的并行特点,提出了一种并行 PSO 优化神经网络算法。最后通过实验比较可以看出,基于 Spark 框架的 PSO 优化神经网络算法的准确度高,且相较于传统基于 Hadoop 的预测方法在处理速度上有显著提高。

关键词 Spark 计算框架,粒子群算法,并行 PSO 优化神经网络,电力通信网络,安全态势预测
中图法分类号 TP391 **文献标识码** A

Prediction about Network Security Situation of Electric Power Telecommunication Based on Spark Framework and PSO Algorithm

JIN Xin[1] LI Long-wei[1] SU Guo-hua[2] LIU Xiao-lei[2] JI Jia-nan[3]
(School of Information, Central University of Finance and Economics, Beijing 100081, China)[1]
(Beijing State Power Communication Network Technology Company, Beijing 100070, China)[2]
(Personnel Testing Center, Ministry of Human Resources and Social Security, Beijing 100011, China)[3]

Abstract With the expansion of the scale of electric power communication network, the electric power communication network continuously produce huge amounts of data communication. At the same time, the communication network attack means is in constant evolution, which brings threats for the safety of the electric power communication network. To solve above problems, combining with the Spark big data computing framework and the advantages of PSO, the Spark memory computing framework of parallel PSO optimization neural network algorithm is put forward to predict the security situation of electric power communication network. This study first introduced the Spark computing framework, the Spark frame has the characteristics of memory computing and quasi real-time processing, accord with the requirement of electric power communication big data processing. Then PSO optimization algorithm was proposed to modify the weights of neural network, in order to increase the study efficiency and accuracy of neural network. Then with the combination of RDD parallel characteristic, this paper proposed a parallel PSO optimization neural network algorithm. Through experiment and comparison, you can see that Spark framework based PSO optimization neural network algorithm has high accuracy, and compared with prediction method based on Hadoop, its processing speed has improved significantly.

Keywords Spark computing framework, Particle swarm optimization, Parallel optimization PSO neural network, Power communication network, Security situation prediction

随着电力通信网络规模的不断拓展,各种网络威胁也日渐增多,设计高效、可靠的安全态势感知(Security Situation Awareness, SSA)方法成为近年电力企业通信安全工作的主要任务之一[1]。SSA 的工作原理是通过收集当前时刻网络中所有关键节点的行为信息,来检测是否存在违反安全策略的行为及被攻击的迹象,并加以统计来描述当前安全态势[2]。

在网络安全态势研究之初,网络环境在规模、速度和应用的多样性方面都无法与今天相比,对网络安全态势的研究也主要以中、小规模为主,选择的数据源多是各类日志记录[3]。Tim Bass 最早提出网络态势的概念,并在文献[4]中提出基于多传感器数据融合的入侵检测概念框架。目前,劳伦斯伯克利国家实验室开发了"Spring Cube Of Potential Doom"系

本文受国网科技部项目(SGTYHT/14-JS-188)资助。

金 鑫(1974-),男,博士,教授,主要研究方向为商务智能、大数据技术,E-mail:jinxin@cufe.edu.cn;李龙威(1992-),男,硕士生,主要研究方向为大数据技术,E-mail:18810539367@163.com(通信作者);苏国华(1982-),男,助理工程师,主要研究方向为电力云计算;刘晓蕾(1983-),女,硕士,工程师,主要研究方向为信息安全;季佳男(1985-),女,硕士,工程师,主要研究方向为信息管理与数据挖掘。

统[5]，实现了流量三维空间视图。美国国家高级安全系统研究中心开发了 SIFT 工具集[6]，以完成全网的网络安全态势感知。国内在网络安全态势预测方面，西安交通大学实现了基于 IDS 和防火墙的集成化网络安全监测平台[7]。在大规模网络安全预测方面，国防科技大学的胡华平等人提出了大规模网络预警系统的基本框架[8]。上海交通大学以 RBF 实现态势安全值的预测[9]。哈尔滨工程大学以 GA-BPNN 神经网络的方式实现态势安全值的预测[10]。文献[11]基于 GRNN-PSO 算法对网络安全态势进行预测，其优点是在中小型规模网络中预测精度高，收敛速度快，但对大型规模网络效果不理想[12]。文献[13]对 PSO 算法收敛于局部最优解进行了改进，并将其应用于解决背包问题，取得一定效果。

以上方法和理论具有很多参考借鉴的方面，但是随着电力通信网络规模的不断扩大，现在的电网具有数据量大、实时性要求高等特点，仅仅依靠原有的防火墙、入侵检测、防病毒等单一网络安全防护技术已不足以应对电网安全的需求[14]。结合电力通信网络的新特点以及大数据技术的发展，本文拟采用 Spark 框架作为内存计算框架，同时应用粒子群算法(Particle Swarm Optimization，PSO)优化神经网络算法对网络安全态势进行预测。

本文第 1 节主要介绍 Spark 系统框架设计，包括数据来源分析、有效数据提取和 Spark 平台处理数据的策略等，为后续的 PSO 优化神经网络算法的预测工作提供了最有价值的数据源；第 2 节主要介绍 PSO 优化神经网络算法以及对其进行的改进；第 3 节通过实验展示了结合 Spark 平台和 PSO 优化神经网络算法在电力网络安全态势预测上的优势；最后总结全文。

1　电力通信网络安全势态大数据计算框架设计

1.1　电力通信网络安全态势概述

目前，电力通信网络正朝着大规模、分布式的方向发展，网络的攻击行为也向着复杂化、规模化的方向发展。为了保证电力通信网络的安全运营，必须对网络安全态势进行感知和评估，从而对攻击行为做出实时响应。

电力通信网络安全态势是指由各种电力通信网络设备的网络行为、运行状况和用户操作等因素构成的整个电力通信网络运行的实时状态和衍化趋势。因此，安全态势是一个全局的概念，是网络运行的状态和趋势。在大规模的电力通信网络环境中，对可能引起网络安全态势的要素进行获取、理解，对要素进行处理并加以分析，可以显示以及预测电力通信网络安全状况的发展趋势。决策者借助电力网络安全态势感知工具显示当前网络环境的连续变化状况，准确地做出决策。

为了预测电力通信网络安全势态，我们需要对体现网络运行状况的安全势态指标进行计算，将海量的网络安全信息归并融合成一个或一组有意义的数值，该数值称为网络安全势态值，管理员通过该数值的变化可以判断网络是否受到威胁或攻击。本文结合 PSO 算法和神经网络，设计了一个基于 Spark 机制的大数据计算框架，基于海量数据计算预测网络安全势态。

1.2　基于 Spark 的系统框架设计

随着电力通信网络的延伸和通信网络攻击手段的增加，电力通信网络的数据来源较多，数据类型繁多。必须对其进行数据采集和提取才可以为电力通信网络安全态势预测提供有价值的数据，然后通过大数据分析技术计算电网安全势态。本研究提出了基于 Spark 的面向电网安全势态分析的大数据数据处理框架，为本文后续提出的电网安全态势预测算法的实施提供了平台依托。整体框架如图 1 所示。

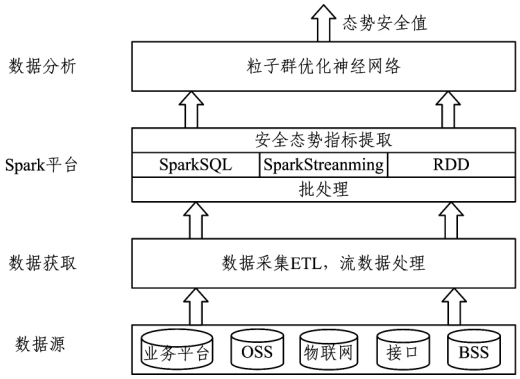

图 1　电网安全态势预测大数据计算框架

整个系统框架包括 4 层：数据源层汇集各种多源海量数据；数据获取层对多源异构数据进行数据采集和预处理；Spark 平台提供内存式实时处理并行计算机制，计算安全势态的相关指标；数据分析层通过算法分析电网安全势态。

1.3　电力通信网络的数据源和数据获取

对于大规模的电力通信网络而言，1)网络节点多、分支复杂、终端类型复杂、数据流量大；2)电力通信网络攻击技术呈多渠道、多手段和自动化的发展趋势，网络攻击的隐蔽性和潜伏时间也越来越复杂，电力通信网络威胁不断增多。电力通信网络数据主要来源于电网调度信息、电力公司基层业务信息、各站点之间的通信信息包括移动终端设备和视频会议等。这些数据分散存储于各个平台，如：业务平台、互联网平台、运营支持系统(Operation Support System，OSS)等。通过抽取转换加载(Extract-Transform-Load，ETL)技术将数据从这些平台中抽取出来，经过数据清洗，将预先定义的与电力通信网络安全态势评估有关的数据加载到数据仓库中。

1.4　电力通信网络数据处理策略

传统的电力大数据安全态势预测系统普遍采用 Hadoop 框架，但 Hadoop 框架不能有效处理频繁迭代计算等问题[15]；同时 Hadoop 框架的 MapReduce 的架构设计使得它在执行任务时，各个任务间的状态共享和全局同步非常困难，只能通过 HDFS 完成交互。相较于 Hadoop，Spark 基于内存计算框架，将一份待处理的数据缓存到内存，然后在同一份数据上进行迭代计算，因此对于迭代次数较多的数据处理方法应选用 Spark。

Spark 可以比其他平台更高效地预测网络安全态势的原因是它的弹性分布式数据集(RDD)。RDD 本质上是内存数据集，其对内存的读写速度远远超过对磁盘的读取速度，可以大大减少 I/O 操作的时间。另外 RDD 支持容错，通常支持容错的方法主要有两种：复制数据和日志。由于复制数据的容错机制存在数据冗余以及会增加电力通信网络通信的开销，因此 Spark 系统采用日志数据更新的方式进行容错[16]。

电力通信系统不间断地产生大量的信息，对实时数据处理能够更好更快地预测出通信网络中的危险态势。Spark Streaming 是构建在 Spark 基础上的实时计算框架。Spark Streaming 将输入的任务流拆分为一系列秒级的批处理作业，

并交由 Spark 引擎处理。每一段数据都转换成 Spark 中的 RDD，然后将 Spark Streaming 中对 DStream 的操作变为针对 Spark 中对 RDD 的 Transformation 操作[17]，Spark Streaming 可以读取 HDFS、Flume 等平台中的流数据。最小窗口可以选择在 0.5～2s 之间，这已经可以满足电力大数据安全态势预测的准实时计算场景。

2 基于 PSO 优化神经网络算法的网络安全态势预测

为了预测电力通信的网络安全势态，我们需要在大数据环境下计算网络安全势态的指标。传统的神经网络方法存在局部极小化和收敛速度慢等问题，文中结合 PSO 算法和神经网络，提出用于电力通信网络安全态势预测的 PSO 优化神经网络算法，并结合 Spark 的 RDD 迭代运算特点对 PSO 优化神经网络算法做了并行化改进。针对电力通信数据，运用基于 Spark 的并行 PSO 优化算法对网络进行安全态势感知，对电网安全态势进行评估和预测，与基于 Map Reduce 框架的方法相比，所提方法具有良好的实时性和扩展性。

2.1 前馈神经网络和 PSO 算法的建模过程

神经网络是模仿人类神经结构和功能的信息处理系统。前馈神经网络是应用较多的神经网络系统，典型的前馈神经网络有 BP 神经网络等。BP 神经网络通过误差反向传播调整网络参数，BP 神经网络算法的基本思想如下：首先通过正向传播，输入层接受样本值，经过各层神经网络计算之后输出结果，如果输出值与期望值之差超出允许范围，则进行误差反向传播，将误差通过隐含层逐层传递到输入层，同时将误差分担给每个神经元，通过获得的误差信号来修正各单元的权值。通过大量重复以上过程来完成神经网络的学习过程，直到达到终止条件。计算网络安全态势值的神经网络结构如图 2 所示。

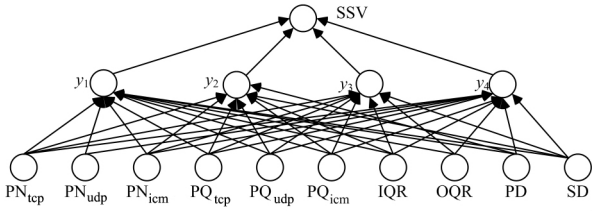

图 2 神经网络预测模型

通常情况下，由输入层、隐含层和输出层构成的 3 层神经网络结构可以被用来回归所有的系统。本文采用具有类似层次结构的神经网络来预测网络安全态势值。参照国家信息安全中心的 Netflow 数据采集展示系统，本文选取数据集中的 10 个字段作为态势指标，并将其作为神经网络预测的输入量，记为 $x_k=(x_{k1},x_{k2},\cdots,x_{kh},\cdots,x_{kn})$。

在设计隐含层的结构的过程中，有经验公式 $m=\sqrt{n+l}+a$。其中 m 表示隐含层节点个数，n 表示输入层节点数，l 表示输出层节点数，a 是 1～10 之间的常数。为了提高准确性，分别试算了 3，4，5，最终确定隐含层的节点数为 4 个。隐含层和输出层的输出采用双极性 S 型函数。

$$f(x)=\frac{2}{1+e^{-x}}-1=\frac{1-e^{-x}}{1+e^{-x}}$$

输入层到隐层的权值矩阵用 V 表示，隐层到输入层的权值矩阵用 W 表示，对于隐层，有：

$$y_j=f(net_j),j=1,2,\cdots,w$$

$$net_j=\sum_{i=0}^{n}v_{ij}x_i,j=1,2,\cdots,m$$

对于输出层，有：

$$O_k=f(net_k),k=1,2,\cdots,m$$

$$net_j=\sum_{i=0}^{n}w_{jk}y_i,j=1,2,\cdots,w$$

因为本文的输出只有一个，即网络安全态势值，所以输出层节点为 1 个。每次训练之后与真实值进行比较，如果得不到期望输出，则进行反向传播，直到验证误差减小到允许误差的范围内。

PSO 算法是近年来的研究热点，它是由群居动物的群居行为启示得到的。PSO 算法的基本思想是：将群体中的个体看作是多维空间的一个没有体积和质量的粒子，每个粒子有一个二维初速度，每个粒子根据自身飞行经验和群体飞行经验对自己的飞行状态进行修正，即每个粒子通过自身的历史最优解和群体的最优解来修正自己飞行的速度和方向，从而形成群体寻找最优解的正反馈机制。

设第 i 个粒子的位置 Y_i 和速度分别为：

$$Y_i=(Y_{i1},Y_{i2},\cdots,Y_{in})$$

$$V_i=(V_{i1},V_{i2},\cdots,V_{in})$$

第 i 个粒子的历史最好点为：

$$P_i=(P_{i1},P_{i2},\cdots,P_{in})$$

粒子群粒子最好点为：

$$P_m=(P_{m1},P_{m2},\cdots,P_{mn})$$

粒子的速度和位置按如下方程进行更新：

$$V_{in}(t+1)=wv_{in}(t)+c_1r_1(P_{in}-x_{id}(t))+c_2r_2(P_{mn}-x_{in}(t))$$

$$Y_{in}(t+1)=Y_{in}(t)+V_{in}(t+1)$$

其中，w 是惯性因子，取值范围是 0.4～0.9；c_1 和 c_2 是学习因子，通常值为 2；r_1 和 r_2 是在[0，1]内均匀分布的随机数。

2.2 PSO 优化神经网络算法

通过对 PSO 和神经网络算法的分析，结合近年来的研究热点，采用群体智能优化神经网络算法来预测网络安全态势。近年来随着研究的日益深入，许多研究者利用群智能优化算法来优化神经网络的训练过程[18]。PSO 算法具有较好的全局收敛性和鲁棒性等特点，因此本文利用 PSO 算法来优化神经网络算法。经过 PSO 优化的神经网络算法既可以提高神经网络的学习能力及收敛速度，又可以利用神经网络简单的学习规则和非线性映射能力。

为了解决神经网络局部极小化和早起收敛速度慢等问题，本文提出利用 PSO 算法来优化神经网络，其基本思想是利用 PSO 算法替换神经网络的误差修正算法。粒子群中的每个粒子代表一组向量，对应神经网络中的一组参数。这样寻找神经网络最优参数的过程就转化为寻找粒子群最优解的过程。当达到终止条件时，程序终止，得到的最优解就是神经网络模型的最优参数。

PSO 优化神经网络的伪码如下。

输入：训练集，测试集，最大迭代次数

输出：PSO 优化神经网络模型，性能评估表

```
BEGIN
/*初始化粒子群 */
```

```
for Index=1:dimSize
  postion=x(Index)+x(dimsize+Index);
  if postion>particle_dimension_maxValue
    postion=particle_dimension_maxValue;
  else if postion<particle_dimension_minValue
    postion=particle_dimension_minValue;
    end
    x(Index)=postion;
end
for iter=1:iterMax
  for particleIndex=1:particleIndexSize;
    let Result=
    Partfit<objFun(pBest)&&cycInd<MaxInd;
    while(Result is Ture)
    /*由 PSO 更新神经网络权值*/
      let(y0,y1,…,yn-1)=updating(x0,x1,…,xn-1);
      let Result=Particle_fitness(y0,y1,…,yn-1);
      continue;
  end
end
/*检测神经网络模型 */
let Model=get_model(Result);
let train_set=training_set(t1,t2,…,ti);
let Result=cycle_Index>cycle_Index_max;
while(Result is FALSE)
  let (t1',t2',…,ti')=compute(Model,t1,t2,…,ti);
if(Result is TURE)
  return(Model,Performace_Evaluation_results);
else
  return 0;
END
```

2.3　结合 RDD 的并行 PSO 优化神经网络

为了尽可能提高电力通信网络安全态势预测的效率，不仅要利用 Spark 框架的 RDD 提供的计算接口并行化效果，还需要对算法进行并行化优化。设计和实现算法时，在实践中需要重点考虑 RDD 分区的划分、shuffle 操作的时机与次数，实现良好的 Spark 算法与糟糕的算法在性能上可以导致上 10～100 倍的差别[19]；同时，传统粒子群优化算法在进化过程中具有易收敛到局部极值从而出现早熟的现象，进化后期存在收敛速度慢[20]等缺陷。为了解决以上问题，同时为了提高电力通信网络安全态势预测的效率，本文对 PSO-神经网络模型进行了并行化优化。

将粒子群体随机分解为 N 个子种群，各子种群以相应的搜索方向更新粒子位置的进化方法，采用二值交叉算子将每个子群中最差的 $N-1$ 个位置替换成其他 $N-1$ 个子群的最优位置。

具体算法步骤如下：

步骤 1　对电力态势指标训练数据进行预处理，将训练集分解成多个子集，存储在 HDFS 中。

步骤 2　采用 PSO 算法对神经网络的权值进行全局寻优，得到优化的网络初始权值；其中所述步骤 2 进一步包括步骤 2.1－2.7。

步骤 2.1　在取值范围内随机生成粒子群，包括粒子初始速度和初始位置。

步骤 2.2　将粒子群随机分成 N 个相等的子群体，生成粒子群 RDD 数据集；设置迭代 N 代后每个子群中最差的 $N-1$ 个位置替换为其他 $N-1$ 个子群的最优位置。

步骤 2.2　启动 Spark 集群中的 map 接口的作业。

步骤 2.3　将步骤 2.2 生成的粒子群 RDD 作 map 转换处理，训练神经网络，由适应度函数计算各个粒子的适应度值。

步骤 2.4　比较各个粒子的适应度值和历史最优值，如果当前的适应度值更优，则更新历史最优值。

步骤 2.5　比较各个粒子的历史最优值与该子种群的最优值，如果某个粒子的历史最优值比该子种群的最优值更优，更新该子种群的最优值。

步骤 2.6　根据相应的搜索方向更新粒子的位置，生成新的粒子群 RDD。

步骤 2.7　判断是否满足终止条件，如果满足则输出，否则进入步骤 2.2 进行下一轮迭代。

步骤 3　在各计算节点上均采用优化后的网络初始权值建立自身的神经网络结构。

步骤 4　采用电力态势指标训练集对 BP 神经网络算法进行迭代，包括以下步骤：

步骤 4.1　从 Driver 进程得到神经网络参数初始权值，广播到每个 Slave 节点。

步骤 4.2　根据初始权值，在每个 Slave 节点上实例化一个神经网络结构。

步骤 4.3　根据每个 Slave 节点得到的部分电力态势指标训练样本训练神经网络，达到预定的迭代次数或者训练误差达到精度要求范围之内后得到神经网络权值。

步骤 4.4　由每个 Slave 节点的参数权值和参数权值的调整量决定是否需要再次迭代。

步骤 5　输出神经网络结构。

本文采用 Yarn 方式作为多 Slave 节点的调节方式，对应的 Master 节点和各 Slave 节点的算法流程如图 3 所示。

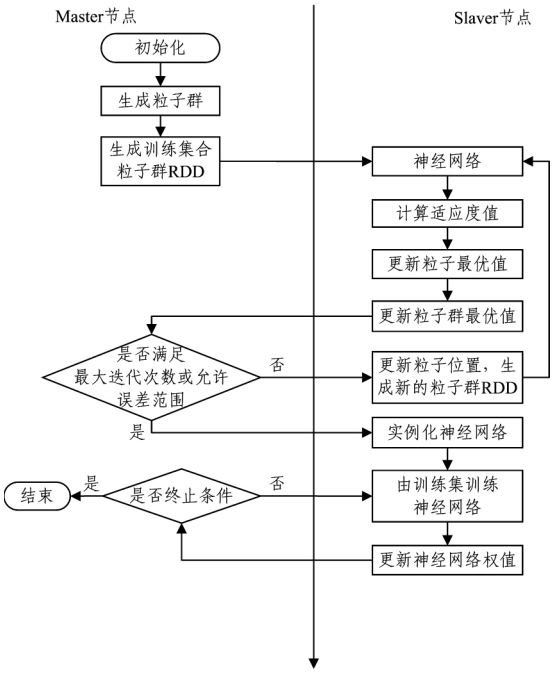

图 3　并行 PSO 优化神经网络算法的流程

3 实验分析

实验主要针对两部分做对比分析：1）并行 PSO 优化神经网络算法预测网络安全态势的预测分析；2）基于 Spark 和 Hadoop 的 PSO 预测网络安全态势的对比分析。

实验使用平台配置：ThinkServer RD650，处理器类型是 6 核 Intel® Xeon® CPU E5-2620@2.40GHz，内存 64GB，硬盘 16T。虚拟机使用的是 64 位版本 Ubuntu14.10，使用的 Spark 版本是 1.2.0。数据集来自国家电网近 4 年的电力通信网络数据，大约有 1.7G。

3.1 并行 PSO 优化神经网络算法的网络安全势态预测

本实验采用电网公司提供的数据集，数据集中包括通信业务基础数据、应用系统基础数据和外部数据以及真实的网络安全态势值。真实的网络安全态势值是由安全管理中心通过入侵检测、一段时间的综合日志分析、用户提交的安全事件报告和安全产品开发报告商报告获得的网络安全状况评价的综合评分，这个数据是真实情况的反馈，不是由计算得来的。本实验只选取与通信网络安全态势预测有关的部分，经过数据的预处理得出电力通信网络数据包的统计信息。抽取统计信息中本系统输入的 10 个态势指标对模型进行训练，训练完成之后输入待测数据，计算出相应的态势安全值（SSV）并将其与真实的网络态势安全值进行比较，结果如图 4 所示。

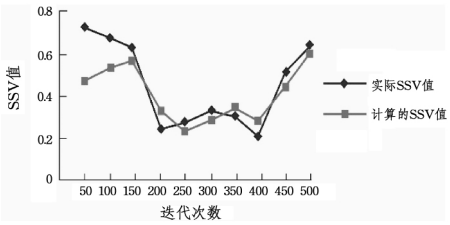

图 4 SSV 计算结果对比图

从图 4 可以看出，由并行 PSO 优化神经网络得出的网络态势安全值与实际值的误差在一定范围内波动。当迭代次数较少时，准确率较低，在 70%左右；而当迭代次数增大之后准确率会有显著提高，可以达到 95%左右。根据计算的势态安全值参照表 1 中的安全势态对应标准，将危险级别划分为 5 类。针对不同级别的危险，可以设置不同级别的预警手段，及时通知电力企业相关部门，采取相应的措施来处理预测到的危险态势，达到预测和预警网络安全态势的目的。例如：当预测的安全态势值在[−0.25，0.25]之间时，系统判断危险级别为一般危险，将自动触发警铃报警，及时提醒电力企业员工将要发生网络攻击，员工可以提前做出应对措施。

表 1 态势安全值对应表

态势安全值	危险级别
[1，0.625)	没有危险
[0.625，0.25)	轻度危险
[0.25，−0.25)	一般危险
[−0.25，−0.625)	严重危险
[−0.625，−1)	特严重危险

3.2 Spark 框架和 Hadoop 框架下的实验对比分析

为了体现 Spark 在计算性能上的优势，本文选用 1.7G 的电力通信网络数据，分别将数据运行在 Spark 和 Hadoop 计算平台上，Spark 平台上运行的是并行化的 PSO 优化神经网络算法，Hadoop 平台运行的是 PSO 优化神经网络算法。将两次结果进行对比，结果如图 5 所示。

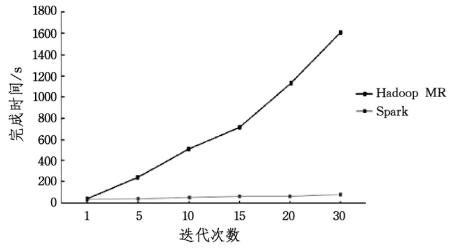

图 5 PSO 优化神经网络不同迭代次数的性能对比

如图 5 所示，结合并行 PSO 优化神经网络的 Spark 计算模型比 Hadoop 平台在性能上有很大提升，尤其是当迭代次数增加时，性能提升尤其明显。这是因为 Spark 是基于内存计算，中间结果放在内存中，并且采用了并行的 PSO 优化神经网络，极大地提升了 Spark 并行计算的能力，减少了迭代作业的完成时间。

另外，PSO 优化神经网络算法的实现还受到神经网络节点数的影响，下面分别取样本数为 1 千万条的样本集，运行 Spark 和 Hadoop 上对应版本的 PSO 优化神经网络算法，对比结果如图 6 所示。

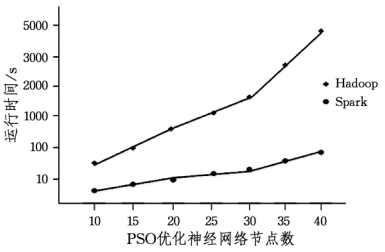

图 6 不同神经网络节点 Hadoop 与 Spark 的性能对比

从图 6 可以看出，随着 PSO 优化神经网络节点数的增加，系统的计算复杂度增加，Spark 和 Hadoop 的运行时间也会相应增加，但 Spark 的运行效率比 Hadoop 平台高出 50 倍。

通过调整迭代次数和 PSO 神经网络节点数，经过多次验证，Spark 结合并行 PSO 优化神经网络算法的计算效率比 Hadoop 平台高出 50 到 100 多倍。实验结果表明：利用 Spark 内存计算框架的较高的计算性能，采用并行 PSO 优化神经网络算法对电力通信网络进行安全态势评估具有预测速度快、准确性高等特点。

结束语 通过对基于 Spark 框架的并行 PSO 优化算法的网络安全态势预测的研究与实现，可以发现该模型能够结合 Spark 框架和神经网络的优点，对电力网络安全态势具有非常好的预测作用。因此，采用基于 Spark 框架的 PSO 优化算法对电力网络安全态势进行预测具有重要意义，可为电力网络安全人员防范电力网络攻击提供可靠支持。

参 考 文 献

[1] 徐茹枝，王宇飞，等. 面向电力信息网络的安全态势感知研究[J]. 电网技术学报，2013，37(1)：128-134.

[2] SPEROTTO A，SCHAFFRATH G，SADRE R，et al. An overview of IP flow-based intrusion detection[J]. IEEE Communications Surverys& Tutorials，2010，12(3)：343-356.

[3] 姚东. 基于流的大规模网络安全态势感知关键技术研究[D]. 郑州：中国人民解放军信息工程大学，2013.

[4] TIM B. Intrusion systems and multisensor data fusion：creating cyberspace situational awareness[J]. Communications of the

ACM，2000，43(4)：99-105.
[5] LAU S. The spinning cube of potential doom[J]. Communications of the ACM，2004，47(6)：25-26.
[6] YURCIK B. Security Incident Fusion Tool (SIFT) Research Project[OL]. http://www.projects.ncassr.org/sift.
[7] 陈秀真，郑庆华，管晓宏，等. 网络化系统安全态势评估的研究[J]. 西安交通大学学报，2004，38(4)：404-408.
[8] 胡华平，张怡，陈海涛，等. 面向大规模网络的入侵检测与预警系统研究[J]. 国防科技大学学报，2003，25(1)：21-25.
[9] 任伟，蒋兴浩，锬锋. 基于 RBF 神经网络的网络安全态势预测方法[J]. 计算机工程与应用，2006，42(31)：136-138.
[10] 胡明明，王慧强，赖积保. 一种基于 GA-BPNN 的网络安全态势预测方法[EB/OL]. [2007-07-24]. http://www.paper.edu.cn/releasepaper/content/20070-437.
[11] 王宇飞，沈红岩. 基于改进广义回归神经网络的网络安全态势预测[J]. 华北电力大学学报，2011，38(3)：91-95.
[12] 丁硕，常晓恒，巫庆辉，等. GRNN 与 BPNN 的函数逼近性能对比研究[J]. 现代电子技术，2014，37(7)：114-117.
[13] 涂娟娟. 粒子群优化算法的几种改进算法及应用[D]. 南京：中国矿业大学，2014.
[14] 曹蓉蓉. 大数据环境下网络安全态势感知研究[J]. 数字图书馆论坛，2014(2)：85-91.
[15] 孟建良，刘德超. 一种基于 Spark 和聚类分析的辨识电力系统不良数据新方法[J]. 电力系统保护与控制，2016，44(3)：85-91.
[16] 胡俊，胡贤德，程家兴，等. 基于 Spark 的大数据混合计算模型[J]. 计算机系统应用，2015，24(4)：214-218.
[17] ZAHARIA M，DAS T，LI H，et al. Discretized streams：an efficient and fault-tolerant model for stream processing on large clusters[C]// Proceedings of the 4th USENIX conference on Hot Topics in Cloud Ccomputing. USENIX Association，2012.
[18] 涂娟娟. PSO 优化神经网络算法的研究及其应用[D]. 南京：江苏大学，2013.
[19] 唐振坤. 基于 Spark 的机器学习平台设计与实现[D]. 厦门：厦门大学，2014.
[20] 刘世成，张建华，刘宗岐，等. 并行自适应粒子群算法在电力系统无功优化中的应用[J]. 电网技术，2012，36(1)：108-112.

(上接第 335 页)

从安全和时间方面对其性能进行分析。

5.1　安全性能分析

本文所提出的 HBES 的安全性着重体现在 ECC 对私钥加密方面。由于 ECC 属于非对称加密算法，安全性优于对称加密算法，与其他对称加密算法相比，ECC 的安全性依赖于椭圆曲线上的离散对数问题(ECDLP)的求解困难性。研究表明迄今为止没有发现有效计算 ECDLP 的方法[6]。

可以看出 HBES 具有良好的安全性，因此可较好地适用于云审计中企业数据的加密。

5.2　时间性能分析

云审计数据侧重于对数据的加密上传过程，因此，基于加密耗时来衡量 HBES 的性能。通过比较 ECC 与 HBES 的加密耗时，验证该方法在保证安全性的前提下具有较快的加密速率。

以行为单位读取纯文本文件的方法对字符串进行加密，并把加密内容存储在另一个文件中，计算文件加密时间。加密大文件时，需在虚拟机中运行 32 位 Linux 系统，配置(－Xmx1535M－Xms1536M)设置内存大小。

图 5 反映了两种加密方式对应不同数据量时所花费时间的关系。对比图中数据可知，HBES 的加密时间远远小于 ECC 的加密时间。另外，对于较大数据，ECC 加密时间显著增加，而 HBES 呈平稳式上升。因此，该方法具有较好的加密效率，适用于云审计中大数据量的加密。

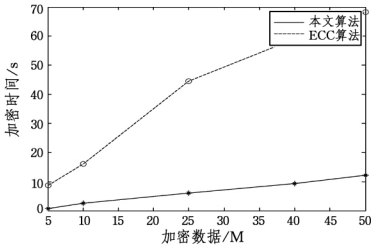

图 5　加密时间对比

结束语　本文对审计数据在网络中传输所带来的安全问题的基础上，引进了混合双向加密方案(HBES)，以保证待审计数据的安全性。HBES 引入了双向加密的概念，较之前的单向加密方法更加突出了从云端下载数据时的加密。提出在上传或下载数据时，由事件发起者产生 ECC 私钥，在网络传输过程中只有公钥和密文可见，而私钥则由发起者保存在本地系统中。当密文在传输过程中被非法窃取，由于私钥的本地化存储，使得密文无法被解密从而保证明文的安全性。通过相关模拟实验验证了 HBES 的可行性。由于实验所选取的是部分待审计数据，相对于真实云审计数据有一定差度，因此加密时间具有一定波动性。本文的方法旨在在使用 HBES 提高加解密速率的同时保证数据在传输过程中的安全性。

参 考 文 献

[1] 徐贵丽. 云审计：机遇、挑战与发展趋势[J]. 中国注册会计师，2014(3)：109-112.
[2] 陈恺. 2016 年 RSA 大会信息安全热点[J]. 信息技术与标准化，2016(3)：15-16.
[3] 秦荣生. 云计算的发展及其对会计、审计的挑战[J]. 当代财经，2013(1)：111-117.
[4] 冯登国，张敏，张妍，等. 云计算安全研究[J]. 软件学报，2011，22(1)：71-83.
[5] KAUFMAN L M. Data security in the world of cloud computing[J]. IEEE Security & Privacy，2009，7(4)：61-64.
[6] REN K，WANG C，WANG Q. Security challenges for the public cloud[J]. IEEE Internet Computing，2012，16(1)：69-73.
[7] TRAPPE W，WASHINGTON L C. Introduction to cryptography with coding theory[M]. Pearson Education India，2006.
[8] 洪澄，张敏，冯登国. 面向云存储的高效动态密文访问控制方法[J]. 通信学报，2011，32(7)：125-132.
[9] TEBAA M，EL HAJJI S，EL GHAZI A. Homomorphic encryption applied to the cloud computing security[C]// Proceedings of the World Congress on Engineering. 2012：4-6.
[10] 宗平，周明. 云计算中的数据安全存储和加密模型的设计[J]. 计算机技术与发展，2013(11)：137-140.
[11] 俞经善，王晶，杨川龙. 基于 ECC 和 AES 相结合的加密系统的实现[J]. 信息技术，2006，30(2)：44-46.
[12] 陈原. 公钥加密与混合加密的可证明安全性研究[D]. 西安：西安电子科技大学，2006.