漏洞的预测能力：漏洞挖掘模型和神经网络模型的比较  
Yazdan Movahedi A，\*米歇尔Cukier一个，伊利尔Gashi b  
一个中心的风险及可靠性，马里兰大学帕克分校，美国b中置软件可靠性，市，伦敦大学，伦敦，U.K  
1.简介  
研究人员利用模型从各种漏洞数据库的数据新漏洞披露趋势。大多数这些研究的目标是找到模型，其中最适合的漏洞披露过程中，并使用该模型，以预测，对于给定的产品中发现的漏洞（冈村等人，2009年的数量;吕和吕，1996; Ozment，2007; Rescorla，2003; Rescorla，2005年1月; Alhazmi和Malaiya，2005）。预测信息披露的频率的漏洞，因为它可以帮助他们与资源分配用于这些产品的供应商以及最终用户非常有用。此外，这种估计也可用于评估可以由保险公司使用与产品相关的风险提供有用的信息（Li等，2003）。  
漏洞发现模型（的VDM），开发基于其历史IOR behav-到预字典未来的软件漏洞。虽然的VDM往往是准确的曲线拟合方面，他们可能不会在预测（约和Malaiya，2014）表现良好。事实上，的VDM往往没有强大到足以采取漏洞披露的非线性特性考虑在内。  
在本文中，我们将介绍基于神经网络预测软件漏洞在30天的时间间隔总数非线性建模方法。我们比较predic-  
神经网络模型（NNM）所使用的VDM 9个的COM monly和灰能力。我们应用的模型有四个众所周知的操作系统（OS）（在Windows，Mac的Cisco IOS（思科相关的OS）和Linux），以及四个著名的网络浏览器相关的漏洞数据（IE浏览器，Safari浏览器，火狐和Chrome）。  
我们的工作提出了以下贡献：  
我们介绍使用NNM与漏洞披露相关的nonlinear-伊蒂埃斯建模的方法;  
我们比较了NNM的能力和九周的VDM在预dicting在8家知名软件30天的时间间隔软件漏洞总数（我们预计在年2016年，2017年，2018年报告的漏洞）;  
我们表明，NNM优于预测精度方面在所有情况下的VDM，和7例提供的绝对平均偏差较小的值。  
本文的其余部分安排如下。第2节去文士相关工作。第3节描述的分析中使用的数据集。第4只列出了我们analy- SIS使用的VDM。第5节介绍了使用NNM建模软件漏洞的发现过程的方法。第6节礼物  
使用的数据集与预测模型的结果。第7节讨论的主要结果和一定的局限性。最后，第8节提出了今后工作的结论和规定。  
2.相关工作  
多年来，许多漏洞发现模型（的VDM）已经开发了基于其历史行为来预测未来的软件漏洞。  
在建模软件可靠性最早的努力是一个马氏生灭模型（哈德森，1967年）。的几个软件可靠性增长模型（SRGMs）表征软件缺陷发现的亲塞斯高质量综述被（吕和吕，1996）提供。在造型漏洞发现亲塞斯（安德森，2002）最早研究提出第一VDM称为热力学安德森（AT）模型。 Rescorla（20 03，20 05）提出了一个VDM来估计未发现的漏洞的数量。 Alhazmi和Malaiya（2005）提出SRGMs脆弱的应用探索造型。他们还推出了被称为Alhazmi-Malaiya物流（AML）模型，该模型假定在峰值附近发现率值（Alhazmi和Malaiya，2005）对称形状的logis-抽动VDM。一种基于威布尔分布VDM提出由金（Kim等，2007）。 Li等人。 （Li等人。，2004）经验表明，相比于其它可靠性国防部 - ELS，威布尔模型是在大范围的软件系统，用于缺陷出现更好。

一些研究应用现有的VDM或它们的修改版本，以不同类型的软件包，如操作系统和Web服务器，模拟漏洞的发现率和预字典漏洞的数量可能潜在地存在，但尚未发现（雨等人，2006;和Alhazmi Malaiya，2008; Massacci和Nguyen，2014; Shrivastava等人，2015）。其他统计音译模型试图通过取漏洞数据的偏斜度为consid-关合作，以增加发现漏洞国防部 - 鹅岭的精度（约和Malaiya，2014）或使用聚类技术（Movahedi等人，2018; Movahedi，2019 ），常用于社交媒体研究中使用（Rajabi等，2019）。近年来，一些软件的漏洞披露过程模型，使用传统的时间序列模型，如自回归移动平均（ARIMA）开发（Roumani等，2015）。但是，漏洞披露数据中含有大量非线性的，因此传统的时间序列模型可能不适合（Wang等人，2015年）。 Pokhrel等。 （2017）比较了三个操作系统的线性和非线性的时间序列（即视窗7的Mac OS X和Linux内核）的建模能力。他们开发了基于ARIMA，人工神经网络（ANN）和支持向量机（SVM）设置国防部 - 埃尔斯。  
3.数据集  
在本文中所使用的数据集是从由NIST保持国家漏洞数据库（NVD）收集，并使用相同的方法，接着收集（Movahedi等人，2017）。我们利用该漏洞CVE ID与同在的漏洞等公共repos- itories的日期来比较每个漏洞的报告日期NVD。我们更新了报告的日期以给定的漏洞在任何他们所利用的漏洞数据库的公开已知的最早日期。  
我们将分析与四大知名操作系统相关联的报告的漏洞：的Windows（1995年至2018年），苹果（1997年至2018年），思科IOS（1992年至2018年）（与思科相关的OS）和Linux（1994至2018年） ，以及四个著名的网络浏览器：Internet Explorer中（1997年至2018年），野生动物园（2003至2018年），火狐（2003至2018年）和Chrome（2008-2018）。这些软件已经被选中，因为他们是最广泛使用和在数据库漏洞最多的。图1示出了所有vul-的检测频率  
表格1  
每个软件漏洞的数量。  
每个软件在30天的时间间隔相关nerabilities。我们还绘制了每个软件，以便更好地理解漏洞去tection趋势的180日均线（MOVAVG）。如图所示，发生MOVAVG的所有情况下的最大值2015年对于每一个软件后，我们在这个研究中使用的变量是在30天的时间间隔报告的漏洞的累计数。换句话说，我们将与给定的软件到30天的间隔相关的研究期间，并计数在每个时间间隔检测到的漏洞的总数。  
对于每一个软件，我们分析报道的任何版本的所有漏洞。因此，对于每一个软件，报告任何版本的所有漏洞都包括在内。例如，报告mac\_os，mac\_os\_server，mac\_os\_x和mac\_os\_x\_server的所有漏洞都放在一起打造为Mac漏洞数据库。  
此外，对于我们的分析，我们将与各软件分成两组相关的vulnerabil-两者均数据集;培训和测试。训练数据集包括报告的所有漏洞测试数据集包含在年2016年，2017年报告的漏洞，2016年之前，2018年的表1表示每个软件漏洞总数，以及漏洞数量在火车和测试数据集。  
4.漏洞发现模型（的VDM）  
漏洞发现模型（的VDM）可以分为两类：基于时间的和基于工作量的的VDM。基于时间的VDM计算给定软件的漏洞为日历时间FUNC-重刑，而基于工作量的的VDM，通过Alhazmi和Malaiya（2005）提出，考虑在软件寿命中因环境因素的变化，例如数安装的，特定的软件等的安装基础份额（见（Alhazmi和Malaiya，2005）关于基于工作量的模型的更多信息）。在本文中，我们将只使用基于时间的模型，因为数据来源，我们只使用了有关于漏洞报告日期的数据，而不是不同的软件的安装。图2示出在基于（Massacci和Nguyen，2014）本研究中使用的基于时间的的VDM的分类。这些模型是在文献中最常用的VDM。我们认为，从每个班至少一个模型。

S形的VDM划分漏洞发现的过程分为三个阶段，如图3所示。第1阶段代表学到，ING阶段，从引进的软件开始，一直持续到周期的开始被称为“可持tained增长”，增加了软件的普及（约和Malaiya，2014）的结果。在学习阶段，在漏洞发现强度函数是增函数。第2阶段或线性相位是周期当大部分vulner-能力被期望被检测到。这一阶段的强度函数是恒定的。第3阶段或饱和阶段是当大部分漏洞已经被发现（Massacci和Nguyen，2014）的时期。该vulnerabil-  
培训和测试数据集（见第6节）。  
图2.考虑基于时间的VDM的分类。  
对于饱和度相位两者均发现强度函数是的降低了ING。这一阶段将没有限制，只要漏洞的显著数量仍未被发现出现。  
在基于它们在建模能力的纸中使用的五个S形的VDM扭曲数据可作如下分类：两个右偏态分布（基于伽马VDM，尤尼斯折叠VDM），一个柔性偏分布（基于威布尔VDM ），以及两个对称的分布（Alhazmi-Malaiya物流（AML）模型和正常基于分布的模型）。这些都是的VDM SE-lected，因为他们是在国防部 - 鹅岭漏洞发现过程（约和Malaiya，2014）所使用的最知名的VDM。  
此外，我们还包括四个非S形的VDM：Rescorla指数（RE）模型，Rescorla二次（RQ）模型，NHPP幂律模型，线性模型（LM）。约了将Rescorla模型和线性模型更多Informa的重刑可以在（Rescorla，2005年1月）和（Alhazmi和Malaiya，2006）分别发现。当模拟故障的平均累积数  
（T），用于软件的可靠性评价，从一个非齐次泊松过程（NHPP）衍生模型经常被使用。阿洛迪（阿洛迪，2015）表明，发现的漏洞可以fol-低幂律分布。在本文中所用的模型在涂布于漏洞数据作为VDM（Movahedi等人，2017; Movahedi等人，2018）。该模型的主要假设是发现漏洞的数量遵循nonhomoge- neous泊松过程。此外，在基于NHPP软件重新责任生长模型（SRGMs），强度函数（ω（T）=的dE [（T）] / dt）的被认为是一个单调函数（Yang和郭，1999）。  
所有的讨论模型的方程中提出  
表2。  
5.的神经网络模型（NNM）  
神经网络模型（NNMS）由一组算法用于建模和识别图案的。 NNMS已被广泛用于预测与连续时间序列数据的数据，如一个城市或股票价格的月度电力需求（Wang等人，2015年;。Adebiyi等2014; Bennett等，2013）。不像的VDM，NNMS能够存在于嘈杂时间序列数据的非线性整合。此外，NNMS不依赖于对基本模型形式的假定，因为它们是完全驱动模型数据来构建。换句话说，NNMS是驱动模型具有强大的预测能力flexi- BLE非线性数据。  
驱动模型的数据是非常有用的情况下，如果没有解释数据的生成过程中的任何适当的理论指导。经验已经表明，NNMS是能够预测线性和非线性时间序列的不同形式的（Kourentzes等人，2014）。  
在这项研究中，预测发现vulnerabili-关系得到随着时间的推移对于给定的软件数量，我们采用前馈NNM，这是使用最广泛的神经网络（Wang等人，2015年）。前馈NNMS在时间接受的输入的固定数目，并产生一个输出。我们假设自命FU-漏洞的数量取决于漏洞解散关闭在过去时期（滞后）的数量。  
在这项研究中，我们使用了一个步进提前预测一个隐层NNM。据（Aslanargun等人，2007），赎罪GLE隐藏层NNM能够近似以任意精度任何非线性函数的。图4显示了在我们的研究中使用的NNM的结构。我们的前馈NNM包括三个层次称为输入，隐藏和输出。每一层是神经元（节点）的集合，其中连接由任意一台既受权重的约束。数据已通过输入层供给，然后它们通过一个或多个隐藏层，并且最终Fi的结果由输出层提供。  
为了预测的现值，使用几个过去的观察。换句话说，该输入是一组的p元件子集

{YT -p。 ..，YT -2，YT -1};和y t是输出或在周期t报告的漏洞的总数。公式。 （1）和（2）示出了与隐藏层的输入和输出值相关联的换mulas，分别。对于输出层中，输入和输出val- UE被由方程表示。 （3）和4。 T-1  
I J瓦特ジ×Y 1 +βJ（J = 1，。。。，H），（1）I = T -p  
ÝĴ，（2）  
H  
I O，（3）  
I表示输入; y表示输出; p和h分别输入和隐藏层节点，的数量 - 误码率;瓦特ジ表示输入和隐藏层的连接权重;并且w OJ去注意到的隐藏层和输出层的连接权重。隐藏层和输出层的偏置值分别由βJ和αO示出，并且总是在-1和1 F H和F O之间进行与隐藏和输出层，分别相关联的非线性激活函数。作为隐蔽层的激活功能，我们使用双曲正切函数，因为它是最广泛使用的功能（Wang等人。，2015）。  
在设计NNM的初始步骤是确定输入节点（滞后）和隐藏层节点的最优数量。根据文献，也没有系统的方法（Wang等，2015）;识别的节点的适当数量的最常用的方法（输入和隐藏的）是通过基于找到的最小的试验和错误均方误差的列列车─ING数据的子集的（MSE），用于初始测试和验证（侯赛尼等人。，2006）。其次，我们根据所介绍的算法（ADE-BP神经网络）的优化方法（Wang等人，2015年）来识别输入的OPTI-MAL号（滞后）和隐藏节点与每个软件相关的时间序列的数量。在（Wang等人，2015）中，示出的是使用ADE-BPNN精度相对预测提高到基本NNMS，ARIMA模型模型（ARIMA），以及其它混合模型的时间序列数据。我们评估了多达50个隐藏节点的每个时间序列，并选择了最小化MSE隐藏节点的数量。我们开始与来自EVAL-uating与每个时间序列相关的偏自相关函数（PACF）的过程中得到的统计显著滞后。在时间序列分析，PACF给出了线性与它自己的滞后值的时间序列的TiAl标准杆相关和评价（古吉拉特和波特，2009）。然而，我们不能只依靠我们从PACF发现，因为，在这种情况下，输入选择将仅仅基于线性模型的鉴定滞后，而对于使用NNM的目标是捕捉非线性相关，为好。对现有NNMS输入选择方法很好的回顾中提供（May等，2011）。  
在本文中开发的NNM使用Mat-实验室R2018a编程。对于每一个软件，我们通过divid-荷兰国际集团的漏洞数据集分成两组开始了我们的分析;培训和测试 - ING。我们使用相同的数据分两种建模方法（和的VDM NNM）。训练数据集由2016年之前所报道的测试数据集所有vulnera- bilities组成报道年2016年，2017年，2018年NNM训练是一个复杂的非线性优化问题vul- nerabilities的。因此，对被困在错误表面局部极小的可能性。为了避免收到效果差，训练过程中应以不同的随机起始重量和偏见反复多次（Kourentzes等，2014）。

我们设置为500个时代的最大训练数量。时代代表的次给定数据集泌尿道感染，lized培训总数和显示的时间在网络中的权重进行了更新（Siami-Namini和那敏，2018）的数量。由于在深学习算法模型优化使用梯度下降法进行（Zhang等人，1998），是有意义的通过网络多次因此通过学习数据集来更新权重，并实现更精确的预测模型（Siami -Namini和那敏，2018）。我们使用列文伯格 - 马夸特（LM）方法作为我们的学习功能。隐藏层和输出层的活化功能是正切S型和purelin功能分别。为了避免过拟合/过训练，对于每个软件，我们通过分割数据集中到训练数据（70％），验证数据（15％），和测试数据（15％）的三个子组中采用的交叉验证方法;并且经由通过Matlab神经网络工具箱提供的指标如梯度下降（梯度阈值= 1.00E-4）和验证检查的最大数目（最大\_fail = 100）检查训练的网络的验证性能。这些指标担任了训练阶段的停止条件。每当在训练网络的参数满足这些阈值，训练过程停止。图5示出了具有用于NNM的Cisco IOS的​​训练阶段相关联的曲线图。由于它显示的基础上，列车数据的误差曲线，没有过度拟合的迹象。而且，随着训练数据的测试子组相关联的曲线图表明，该模型在训练阶段表现良好。 6.结果  
我们用九周的VDM，并在8家知名软件漏洞的发现过程中的一个NNM（4操作系统和  
四个Web浏览器）。所述的VDM拟合使用（Massacci和Nguyen，2014）中描述的非线性回归方法的数据集。  
预测能力的分析开始通过将数据分成两组训练和测试数据。无论是的VDM和NNM使用的数据集，包括报告给定的软件的所有版本的所有漏洞。培训期间，当与给定的软件相关的第一个漏洞被发现，并一直持续到12/31/2015的时间开始。我们计算了预测为年2016年，2017年，2018年和正如示于图1中，暗和亮条分别示出了具有列列车─ING和测试数据集相关联的数据，。我们的漏洞数据分成30天的间隔中的脆弱性analy- SIS文献是共同（约和Malaiya，2014; Alhazmi和Malaiya，2008; Massacci和Nguyen，2014）。  
对于的VDM，在训练期间，训练数据来估计模型参数。为了避免过度拟合，10倍交叉验证还对训练数据进行的。与实际数漏洞进行calcu-后期预测精度进行比较的埃斯蒂配合最终值由九个模型产生的每个时间间隔。对于NNM，对于每一个软件，我们使用的训练数据来训练NNM。用训练NNM，我们在预测PE-荒漠化问题的区间预测下一个值。预测精度是基于所得到的估计和漏洞的实际数目之间的比较。  
对于训练的一部分，为的VDM，我们采用合适的测试（Massacci和Nguyen，2014）的卡方（χ2）善良，评估每个模型如何适合训练数据。在χ2统计量CAL-使用以下等式culated：  
N1（S I - E I）2  
 （5）I = 1个E I  
其中，S i和E I分别是模拟的和预期的观测值在第i个时间点，。 N1是在列车的数据集（用于模拟的时间块）观测值的数目。对于拟合是可接受的，则相应的χ2临界值应比χ2统计量自由的给定α水平和度。我们选择0.05的α水平。零假设表明，实际分布深受拟合模型来描述。因此，如果χ2检验的p值低于0.05，则适合将被认为是不令人满意的。 p值越接近1表示一个更好的配合。对于每个VDM，评估其预测capabil-两者均之前，我们首先检查是否是统计上的声音或没有。如果没有健全的（p值<0.05），我们忽略在我们的评估过程模型。

对于训练的一部分，对于NNM，出有不同数量滞后的训练模型，优化分析模型被选为基于MSE值。最后，对于每个软件中，使用的最佳选择出的分析模型，以便为测试数据集合中的预测（漏洞在2016年，2017年，和2018年报道）。在这项研究中，关于NNMS，我们刚刚报道最好的NNM相关的重新sults，因为它不可能包括每个具有滞后的不同combina-重刑的软件，由于空间限制所有的培训NNMS。我们计算的两张标准化预见性的措施，平均误差（AE）和平均偏差（AB）（约和Malaiya，2014）。 AE是模型的预测效果在整个试验阶段的度量，且AB指示哪个评估其对高估或低估discov- ERED漏洞的数量趋势模型的一般偏压。 AE和AB被定义为：  
AE（6）  
ñ  
= 1  
   
AB（7）  
ñ  
t = 1时  
其中N2是时间点的总数量（每30天）在预测期间，和是总漏洞在时刻t的实际数目，而t是总的漏洞的在时刻t的估计数量。  
此外，为的VDM，我们报告AE我，这表示第i模型的AE，并以最小的AE模式之间的差异百分比。  
AE I =（A E I -Aë分钟）\* 100（8）  
其中AE i是第i个模型的AE，和AE min是所述一组模型检验（即，最佳模型）中的最低AE一个获得。  
均方根误差（RMSE）是经常用于计算拟合误差另一度量。然而，Mentaschi等。 （2013）表明，对于一些应用（例如，实际数据的高波动）RMSE的较低值并不总是模拟的AC- curacy的可靠指标。因此，经校正的估计HH由汉纳和Heinold（1985）中的溶液所构成亲：  
   
= I - O I）2  
HH（9）  
 S I O I  
其中，S i是第i个数据的模拟I，O i是第i个观测（测试数据）i和N2是在测试数据集（用于模拟的时间块）观测值的数目。在接近零HH是，更准确的模型。  
表3-4呈现AE，AB，HH，AE i和p值的值分别（我们用\*显示以P <0.05的模型）为我们每模型AN- alyzed情况（的VDM和NNM）， 。我们还使用了术语“NS”为模式的AE我值和P <0.05，代表不尽如人意。 AB可以是正的（对于高估）或负（对低估），而AE始终为正。在每一种情况下，我们首先通过比较它们的预文辞准确找到了最好的VDM，然后比较那些国防部 - 与NNM ELS的准确性。换句话说，为的VDM，有AE的最小值的模型被选择为具有最佳的预文辞能力和AE值是伴随着它代表了最好的VDM“BV”超级脚本。此外，与AE的的VDM I <2也选为最佳预测的VDM，其中，相比于最好的模型显示出类似的预测能力（模型/与AE模型I = 0）。此外，归一化的ER- ROR值（（T - ）/）;与每种情况下在图绘制6相关联。因为它被示出，以较少的波动模型得到更高的精度。  
基于由表3-4规定的，predic-重刑精度（AE和HH）方面的成果，NNM导致所有我们分析的八个软件的最准确的结果。为了更精确，适用于Windows的NNM的平均误差（AE）为2.6％，并且，至少，比用最好的VDM相关的不良事件，这是电法和RQ较小的1.1％。对于Mac的NNM具有5.3％分别优于最好的VDM（幂律，和RQ），以及较小的6.1％AV-erage错误。对于Cisco IOS，此差值为至少0.1％。 Linux和Chrome的是两个情况下NNM亲比那些从国际志愿组织的VDM更好的预测由是14.9％和24.8％，更准确。对于IE，这区别是至少7.5％。 NNM的Safari浏览器的平均误差比那些最好的VDM（幂律和RQ）小0.3％和1.4％。  
对于Firefox，在NNM 1.2％，1.4％，和1.2％相比，YF，Gamma和威布尔的VDM改善了预测，respec- tively。使用Chrome时，VDM用最小的AE是不是从训练的一部分统计学的声音。所以，我们选择了下一个VDM与p值> 0.05和最小的AE，这是Gamma值。在这种情况下，NNM精度的提高为24.8％。总体而言，在NNM和的VDM之间的预测精度最高differ-分配办法是在铬（24.8％），LINUX（14.9％），IE（7.5％）和Mac（5.3％）分别发现。

另一个因素，它在模式选择位tion的一个重要作用是高估或低估的结果模型的趋势。在这个研究中，我们提供的平均偏差值（AB）以及归一化的预测ER- RORS的视觉波动趋势（图6）。根据表3-4，出八个软件，我们ana-裂解，在误差量方面，NNM跑赢的VDM 7例具有较小| AB |值。只有对于Cisco IOS，由选定的VDM提供偏差的绝对值较小最大为1.3％，比一个起因于NNM。用于Windows，Mac，Linux和IE，Safari浏览器，火狐和Chrome，由NNM提供的偏置MAG-nitudes比那些最好的VDM（在每种情况下较小，我们认为是最好的VDM，其中有最小| AB | ）通过分别为0.7％，5.5％，16.4％，7.5％，0.2％，1.9％，和26.8％。总体而言，在准确性方面，开出8例weanalyzed时，跑赢NNM在所有情况下的VDM。此外，在偏差幅度方面，NNM导致最小偏差值7例。  
7.讨论和限制  
在预测精度（AE和HH），考虑到操作系统和网络浏览器方面，NNM导致比在所有情况下的最佳选择的VDM更准确的结果。如果仅考虑的VDM，幂律VDM被选为最佳模式在以下四种情况进行了8个。伽玛和RQ的VDM各自最好的三种情况下，其他车型出八相比较。韦伯，RE，和YF的VDM各自最好在两种情况下其他车型出来，我们分析8箱子比较。  
在偏见的总体规模（即AB的绝对值）而言，出于我们分析了8起案件中，NNM 7例相比，最好的VDM提供偏置的绝对值较小。只为Cisco IOS，偏压由伽马VDM（0.001）中提供的绝对值小于所述一个从NN导致  
（-0.014）。  
我们认为，最终决定，在同等条件下的精度，在偏置而言，高达研究者选择基于他/她的优先级的最佳模式。但是，从安全角度考虑，最好是选择一个模式，它提供了更多的conser- vative的预测结果。在被选为最好的预测，目前的研究中，模型中，只有两个NNMS（Linux和Safari浏览器）提供高估的结果。其他选择NNMS低估漏洞的数量。它也可以容易地从图6，其中用于Linux和Safari大多数与NNMS相关联的预措辞点都位于所述X = 0轴的推断。  
在两种情况下（即，Linux和铬）时，NNM的预测精度和最好相对VDM之间的差异是非常高的（即，分别为14.9％和24.8％）。我们认为，背后的差异，这些高值的原因是不是与噪音有关，因为我们在数据照顾的噪音问题通过使用交叉验证技术。这是可以假设，对于这三种情况下，NNM能够检测另一个播放 - 荷兰国际集团因子，所述分析的VDM不是相对漏洞发现过程（Iten的等人，2018）。  
总体而言，我们认为，NNM的性能更好的COM缩减到的VDM来自NNM的predict-荷兰国际集团的漏洞披露时间SE-里斯的非线性性质的能力。此外，大多数的VDM考虑漏洞发现过程作为一个纯粹的S形曲线或与漏洞的恒定总数量的单调函数的强度的函数。然而，与给定的软件可以在新的版本改变相关漏洞的数量被释放。此外，的VDM与传统时间序列函数只使用一组用于估计PA-rameters的。在另一方面，由于具有多层感知结构，每层具有多个神经元，并使用不同的一组神经元每NNMS参数提供用于预测的更复杂的结构。当然，我们使用，以避免被困在学习阶段过度拟合特定valida-重刑的方法是使用NNMS的另一个优势。

还有我们的工作一些限制，导致我们无法做出更一般的结论。主要限制是重新GARD使用报告公布的漏洞作为其检测日期的日期。漏洞通常会得到由马licious用户比他们正式公布的时间早检测。为了确保这个估计是尽可能接近的漏洞着称于世的实际日期，我们看到了不同的脆弱性来源和选择了报告漏洞的最早日期。更好的估计可如果我们有计算攻击者的努力和更精确的时候，发现漏洞并报告（例如，在黑暗中网）更多accu-率代理中获得的，而当它在一个公共vul-报道比nerability数据库。然而，获得该数据难以：在黑暗中的网络数据是高度结构化和非常困难的意义添加到什么是开采。另一个限制是对于我们合并报告给定的软件的所有版本有用于训练模型足够的数据全部vul- nerabilities的方式。有适用的VDM到相关的软件不同版本（如Mac OS X中）的漏洞论文（Massacci和Nguyen 2014; Nguyen等，2016）以及该考虑在一起的软件的所有版本研究（ JOH和Malaiya，2014; Movahedi等人，2018）。第一组假定每个软件是一个独立的和明确的产品，但在检测漏洞数据依赖性的来源不是一件容易的任务。  
NNMS，不像分析模型，在数学上并不容易处理，并且不能被人容易理解的。然而，他们能够检测可能由分析模型被错过的机制。因此，它们可以被用来引导用于修改那些模型。在这项研究中，我们发现，更精确的预测还可以通过使用NNMS（Iten的等，2018）。  
的VDM假设故障间隔时间表示产品的总使用时间。我们使用的是日历时间，这可能不是一个使用很好地代表。最重要的是在安全不同的是在估计“攻击者的努力”的难度 - 的攻击者在寻找vulnera-吴春明花费的总时间 - 这是没有必要的可靠性的东西（我们假设用户不小心遇到故障该导致失败，因此使用时间是故障间隔时间足够好的代理）。这方面的一个有用的讨论中的Littlewood等人给出。 （1993年1月）。攻击者的努力是一件很困难的第ES timate和量化。我们研究的目的是因此让尽可能好的使用尽可能的公开可用的安全性数据，以帮助决策。但同时要清楚什么我们可以从这个分析结论的局限性。我们可以从我们目前的分析说，最好的“将在NVD了一个跨VAL吨产品X被报告的漏洞总数为y的信心Z”。我们表明，我们可以最大和最常用的操作系统和网络浏览器家族四做这种预测更好地与NNM比的VDM。对于一些决策者这可能是更多的信息，他们可以结合他们从自己的设备有数据，决定何时操作系统和/或网络浏览器，以及SE-安全支持服务供应来应对使用了宝贵的一块与新的漏洞。  
8.结论和未来的工作  
在本文中，我们预测在三年预测期未来漏洞的数量方面比较的九种常见漏洞挖掘模型（的VDM）具有非线性神经网络模型（NNM）的能力。我们应用所提到的模型有四个知名的操作系统和四个著名的网络浏览器相关的漏洞数据。在国防部 - ELS在预测精度和预测偏差方面进行了评估。结果表明，NNM优于在所有情况下的VDM预测精度方面。在偏见的总体规模而言，从我们分析的8起案件中，NNM提供7例偏差的最小绝对值的COM缩减到最佳的VDM。这项研究表明，神经网络是有希望的随着时间的推移pub- lically报告的软件漏洞总数的准确的预测。

对于今后的工作中，我们计划寻找NNMS对比的VDM的预测能力之间的转播服务的差距背后的原因，专门针对这两个软件，我们发现相当大的差异。我们将尝试探讨当前的VDM是否缺少与他们的数学结构中vulnerabil-两者均发现的过程相关联的机制。我们还计划研究使用机器学到，ING算法等非线性模型结构。其中有递归神经网络（RNN）模型，用于预测时间序列，它可以在建模的依赖两点之间的SE-quence表现得比较好NNMS。通常，在NNMS，我们要选择的输入（输入的数）预先的长度。然后，它是不可能得知取决于发生在很久以前的输入功能。这个问题可以通过具有RNN，理论上可从任意很久以前存储的信息来解决。致谢  
这项研究是由美国国家科学基金会奖＃1223634支持，以及英国EPSRC项目D3S（多样性和深度安全防御：概率方法）下拨款协议700692（DiSIEM）和欧盟委员会通过H2020程序。  
参考  
Adebiyi，A .A。，Adewumi，A .O。，亚青，C.K.，ARIMA与人工神经网络模型的股票价格预测2014年的比较。 [应用Mathe-的matics 1-7。  
Alhazmi，O.H.，Malaiya，Y.K.，漏洞挖掘模型在2008年应用  
阿洛迪，L.，2015年漏洞检测的重尾分布。工程安全软件和系统133-148。  
Aslanargun，A.，马马多夫，M.，Yazici，B.，Yolacan，S.，2007年在时间序列比较ARIMA，神经网络和混合模式：游客到来的预测。  
贝内特，C.，斯图尔特，R.A.，比尔，C.D.，2013年ANN基于住宅水的最终用途去普通话预测模型。与应用40专家系统（4），1014至1023年。  
   
   
侯赛尼，H.G.，洛，D.，雷诺兹，K.J.，2006不同的进料的比较向前推进神经网络结构用于ECG信号的诊断。医疗工程师 - ING与物理28（4），372-378。  
ITEN，R.，Metger，T.，Wilming，H.，德尔里奥，L.，＆伦纳，R。（2018）。发现与神经网络的物理概念。预印的arXiv的arXiv：1807.10300。  
荷兰Joh，H.，Malaiya，Y.K.，2014年模拟偏度在漏洞发现：模拟偏度在漏洞挖掘。质量与可靠性工程国际30（8），1445年至1459年。  
Kourentzes，N.，巴罗，D.K.，科龙，顺丰速，2014年神经网络集成商的时间序列预测。与应用41专家系统（9），4235-4244。  
的Littlewood，B.，布罗克赫斯特，S.，芬顿，N.，梅勒，P.，页面，S.，怀特，D.，...，Goll-曼，D.，1993年迈向计算机安全的操作措施。 [计算机安全2的（2-3），211-229。  
Massacci，F.，阮，V.H.，2014年的实证方法来评估Vulnerabil-两者均发现模型。 IEEE交易软件工程40（12），1147- 1162。  
   
Mentaschi，L.，Besio，G.，卡索拉，F.，Mazzino，A.，基于RMSE波2013年问题  
Movahedi，Yazdan，2019年一些指导漏洞解散covery过程的风险评估。马里兰大学  
Movahedi，Yazdan，Cukier，M.，Andongabo，操作系统和网络浏览器的A.，Gashi，一，2019年基于集群的vulner-能力的评估。计算101（2），139-160。  
阮，V.H.，Dashevskyi，S.，Massacci，F.，2016年所述的用于评估受漏洞版本自动方法。实证软件工程21（6），2268年至2297年。  
Ozment，J.A.，2007年发现的漏洞和软件的安全性。凸轮桥大学取自  
Pokhrel，N.R.，罗德里戈，H.，Tsokos，C.P.，2017年网络安全：采用线性和非桌面操作系统的漏洞的时间序列预测模型  
Rescorla，E，2005年是寻找安全漏洞是一个好主意？ IEEE安全和隐私杂志3（1），14-19。  
Roumani，Y.，Nwankpa，J.K.，Roumani，Y.F.，2015年vulnerabili-关系的时间序列建模。计算机与安全51，32-40。  
   
Siami-Namini，S.，＆那敏，A.S。（2018）。预测经济和金融时间序列：ARIMA与LSTM。预印的arXiv的arXiv：1803.06386。  
王，L.，曾庆红，Y.，陈，T.，与时间序列预测自适应差分进化算法2015年BP神经网络。与应用程序42的专家系统（2），855-863。  
杨T.Y.，郭，L.，1999年为nonho- mogeneous泊松过程的叠加贝叶斯计算。加拿大杂志统计27（3），547-556。  
张，G.，埃迪Patuwo，B.，胡，M.Y.，1998年预测与人工神经NET-

Yazdan Movahedi拥有博士学位度在马里兰大学帕克分校可靠性工程方面的受过。他还拥有硕士学位和学士学位的工业工程学位。他的研究兴趣包括EM-ploying统计方法和机器学习tech- niques评估软件的可靠性和建模漏洞数据。米歇尔Cukier是可靠性恩gineering与马里兰大学帕克分校机械工程系共同聘任副教授。他还为高级CY-bersecurity体验学生（ACES）的董事。他的研究涵盖了可靠性和安全性的问题。他-最新重新搜索专注于网络安全的经验量化。 Cukier博士发表在期刊论文70余篇，并在这些领域审阅会议记录。  
伊利尔·Gashi拥有博士学位在软件的可靠性和计算（2007）及工学士（荣誉）（2003年）度市，伦敦大学tively respec-。他于2003年7月，他目前是一名高级讲师加盟岑TRE软件可靠性（CSR）。目前，他是（2018），在调查深度SE-安全多样性和国防的潜在收益两个项目，一个Prin- cipal调查。他的研究重点是基于软件的系统的可靠性和安全性的定量评估。