Abstract

我们提出了一种前馈神经网络的验证方法，其中所有节点都具有分段线性激活函数。这样的网络经常用于深度学习，并且已经证明很难验证现代可满足性模理论（SMT）和整数线性规划（ILP）求解器。

我们方法的出发点是将整个网络行为的全局线性近似添加到验证问题中，这有助于对网络行为进行类似于SMT的推理。我们提出了一种特殊的验证算法，该算法在搜索过程中采用了这种近似方法，在该算法中，它根据部分节点的相位（phase）分配来推断网络中非线性节点的其他节点相位（phase），类似于经典SAT求解中的单位传播。我们还将展示如何根据搜索过程中执行的分析步骤的结果来推断其他冲突条款和安全节点固定装置。在避免碰撞和手写数字识别案例研究中评估了生成的方法

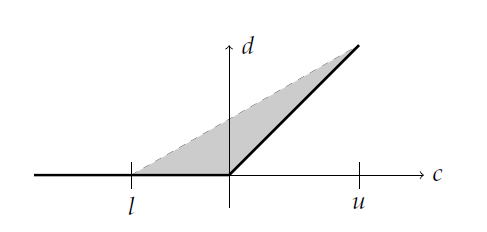
Introduction

在本文中，我们提出了一种针对凸规范验证具有分段线性激活函数的神经网络的方法。该方法支持仅采用分段线性激活功能的现代网络网络体系结构中使用的所有节点类型（例如MaxPool和ReLU节点）。该方法基于可满足性（SAT）解决方案和线性规划的组合，并采用了整个网络行为的新颖线性近似。这种近似使得该方法可以快速排除在验证过程中不考虑节点阶段的较大搜索空间部分。虽然近似值还可以用作SMT解算中的附加约束，并可以改善SMT解算器的计算时间，但我们将其应用到定制的解算器中，该解算器使用来自[CD91]的弹性滤波算法，以在出现以下情况时最小化不可行的线性约束集：冲突，并将其与用于推断隐含节点阶段的专用过程结合起来。这些组件在一起可以大大缩短验证时间。我们将该方法应用于两个案例研究中，即避免碰撞和字符识别，并报告实验结果。我们还将提供最终的求解器和完整的工具链，以使用深度学习框架Caffe [JSD + 14]作为开源软件来生成可验证的模型

Efficient Verification of Feed-forward Neural Networks

定义 relu和maxpool的近似函数

Relu：

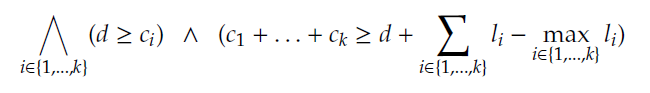


如果relu节点有上下限【l，u】 输入 c 输出d，则近似为



即 图中阴影部分

Maxpool：



最紧密的线性约束（未说明原因）

在建立了一个近似整个网络行为的线性程序之后，我们可以使用它来使所有将来的近似值变得更加紧密。 为此，我们将问题说明作为约束条件添加，并对每个变量v 求解所得的线性程序，同时首先对目标函数1 \*v进行最小化，然后对目标函数􀀀-1\*v进行最小化。 这样就产生了新的更严格的上下限[l， u]对于每个节点（如果网络具有任何ReLU节点），这可用于获得更严格的线性程序。 在流程中包含规范可以使我们获得比没有规范时更严格的界限。 整个过程可以重复几次：每当获得新的上限和下限时，就可以使用它们建立更严格的线性网络近似值，从而可以获得新的更严格的上限和下限。

Conclusion

在本文中，我们提出了一种验证具有分段线性激活函数的前馈神经网络的新方法。 我们的主要思想是生成可以添加到对神经网络验证问题进行编码的SMT或ILP实例中的整个网络行为的线性近似值，并在专门的方法中使用近似值，该方法具有针对神经网络验证的多种附加技术， 它们围绕SAT求解器分组，用于选择网络中的节点相位。 我们考虑了来自不同应用领域的两个案例研究。 该方法允许使用任意凸面验证条件，我们使用它们来定义一个噪声模型，以测试用于识别手写数字的网络的鲁棒性。

尽管本文报告了神经网络验证性能的改进，但在验证方面仍有许多工作要做：我们目前尚未采用专门的启发式方法进行节点相分支选择，而我们的方法却增加了神经网络验证的可扩展性实际上，我们观察到它仍然很脆弱，并且难以验证属性（如MNIST示例中看到的）容易超时。同样，我们必须在实验中简化LeNet架构以进行数字识别，因为原始网络非常大，以至于甚至无法获得网络中单个变量的下限（在开始实际的求解过程之前，我们对所有网络节点都进行了此操作如3.1节中所述，否则将花费30分钟以上，即使这仅意味着求解单个线性程序。而黄等人的方法。 [HKWW17]不受此限制，它无法处理常规的验证属性，这一点很重要。我们计划在将来解决本文提出的方法的网络规模限制。

Introduction

Pulina, L., Tacchella, A.: An abstraction-refinement approach to verification of artificial neural networks. In: Touili, T., Cook, B., Jackson, P. (eds.) CAV 2010. LNCS, vol. 6174, pp. 243–257. Springer, Heidelberg (2010). [https://doi.org/10. 1007/978-3-642-14295-6 24](https://doi.org/10.%201007/978-3-642-14295-6%2024)

Pulina和Tacchella [PT10]为具有非线性激活功能的神经元提供了一种方法，该方法只能扩展到小型网络。在他们的工作中，他们使用具有6个节点的网络，对于大多数实际应用而言，这太少了。他们将反例触发的抽象细化与可满足性模理论（SMT）解决方案相结合

Karsten Scheibler, Leonore Winterer, Ralf Wimmer, and Bernd Becker. Towards verification of artificial neural networks. InMBMVWorkshop 2015, Chemnitz, Germany, pages 30–40, 2015.

Scheibler等。 [SWWB15]考虑具有非线性系统动力学和非线性神经元激活功能的逆摆控制方案的有界模型检查问题，尽管采用了最先进的SMT求解器iSAT3 [SNM + 16]和他们的实验表明，即使扩展此求解器以更好地处理由此产生的问题实例，他们的结果也表明，对于具有26个节点的神经网络，所产生的验证问题已经具有挑战性。

Luca Pulina and Armando Tacchella. Challenging SMT solvers to verify neural networks.

AI Commun., 25(2):117–135, 2012.

Guy Katz, ClarkW. Barrett, David L. Dill, Kyle Julian, and Mykel J. Kochenderfer. Reluplex:

An ecient SMT solver for verifying deep neural networks. In 29th International

Conference on Computer Aided Verification (CAV). Springer, 2017.

Xiaowei Huang, Marta Kwiatkowska, Sen Wang, and Min Wu. Safety verification of

deep neural networks. In 29th International Conference on Computer Aided Verification

(CAV). Springer, 2017.

[HKWW17]描述了这样一种方法，它基于通过网络各层的传播约束。约束对每层输入空间的区域进行编码，这些区域的所有点都导致网络中的总体分类相同。他们的方法部分基于离散化，并专注于鲁棒性测试，即确定输入可以在不更改分类结果的情况下进行更改的程度。它们不支持常规验证属性。

Osbert Bastani, Yani Ioannou, Leonidas Lampropoulos, Dimitrios Vytiniotis, Aditya V.

Nori, and Antonio Criminisi. Measuring neural net robustness with constraints. In

Annual Conference on Neural Information Processing Systems (NIPS), pages 2613–2621,

2016.

[BIL + 16]还以鲁棒性测试为目标，并定义了一个抽象优化约束求解循环，以针对对抗性插管测试网络的鲁棒性。他们还采用他们的方法发现的反例来学习更强大的网络