### 第二十五次2020年6月8日-6月14日

#### 1 验证BNN属性

On Tractable Representations of Binary Neural Networks

二元神经网络的可表示形式

我们考虑将二进制神经网络的决策函数编译成易于处理的表示形式，例如有序二进制决策图（OBDD）和句子决策图（SDD）。首先，我们考虑验证神经网络的鲁棒性的任务，并展示如何在给定OBDD / SDD表示的情况下如何计算神经网络的期望鲁棒性。接下来，我们考虑一种基于伪多项式时间算法的神经网络编译方法。然后，我们在一个手写数字数据集中提供了一个案例研究，重点介绍了从同一数据集中训练的两个神经网络如何具有很高的准确度，但鲁棒性却有很大不同。最后，在实验中，我们表明获得神经网络的紧凑表示SDD是可行的。

第2节介绍相关背景材料。在第3节展示了如何通过将每个神经元编译成布尔电路来将神经网络简化为布尔电路。在第4节讨论如何通过知识汇编获得易处理电路。在第5节中，我们展示了可处理电路如何使人们能够推理神经网络的鲁棒性。我们在第6节中提供了一个案例研究，在第7节中以经验方式评估了我们提出的编译器，最后在第8节中总结。

手机屏幕截图

描述已自动生成

前馈神经网络是有向无环图（DAG）。DAG的根是神经网络输入，称为X1，... Xn。 DAG的叶子是神经网络输出，称为Y1，... Yn。DAG中的每个节点都称为神经元，并包含一个激活函数σ。DAG中的每个边I都附加有权重w。神经网络的权重是其参数，可以从数据中学习。

例如一个神经元的参数对应下列不等式，那么我们可以得到相应的布尔表达式

1.15 · A + 0.95 · B − 1.05 · C ≥ 0.52

=> [¬C ∧ (A ∨ B)] ∨ [C ∧ A ∧ B]

手机屏幕截图

描述已自动生成

知识汇编一直在探讨部分易处理的布尔电路简洁性和易处理性之间的权衡。通过在布尔电路的结构上使用不同的属性，可以以简洁（结果电路的大小）为代价，获得更大的可追溯性（在polytime中执行某些查询和转换的能力）。我们的目标是将神经网络指定的布尔函数编译成易处理的布尔电路，方便解释和验证。

手机屏幕截图

描述已自动生成

图a是神经网络，图b是一个神经元对应的布尔电路，图c是神经网络对应的布尔电路。