隐私保护的分布式机器学习系统原型设计与实现

曾亟

April 1, 2020

Abstract

基于深度神经网络的人工智能近年来取得巨大的成就，但是人工智能技术往往是基于大数据，即大数据驱动的人工智能。但是数据的匮乏往往限制了人工智能的发展，与此同时，数据源之间存在着难以打破的壁垒。一方面，人工智能所需数据来源广泛，遍及各领域，且各领域数据往往是以孤岛形式存在，整合各领域的数据也面临重重阻力。另一方面，随着大数据的发展，世界各国对用户数据隐私和安全管理也日趋严格。因此，要解决大数据的困境，就要保证数据隐私的同时，让机器学习系统高效和准确的利用各自的数据进行学习。我设计了一个分布式机器学习系统的基本模型，在实现模型可利用多方数据信息进行训练的同时，保证了数据隐私安全以及无预测性能损失，并为研究者提供了可拓展的算法接口。实验中，采用当前常用神经网络与数据集测试分布式学习系统的有效性以及模型的性能。

1 Introduction

2 Related Work