

大数据计算基础课程报告

题 目 ： **Plan-Structured Deep Neural Network Models for Query Performance Prediction**

专 业 计算机科学与技术

学 号 2022112249

姓 名 施禹

课 程 高级数据库系统

日 期 2025.5.12

# 摘 要

关键词：查询代价估计

目录

[摘 要 I](#_Toc195993168)

[第1章 题目概述 2](#_Toc195993169)

[1.1 研究问题的背景 2](#_Toc195993170)

[1.2 研究问题的形式化定义与挑战 2](#_Toc195993171)

[1.3 当前研究工作的不足之处 2](#_Toc195993172)

[1.4本次实验要解决的问题以及方法 2](#_Toc195993173)

[1.5 本次实验的贡献 2](#_Toc195993174)

[第2章 基础题的求解 2](#_Toc195993175)

[2.1 求解思路 2](#_Toc195993176)

[2.2 求解过程 2](#_Toc195993177)

[第3章 高级题的求解 2](#_Toc195993178)

[3.1求解思路 2](#_Toc195993179)

[3.2求解过程 2](#_Toc195993180)

[第4章 进阶题的求解 2](#_Toc195993181)

[4.1求解思路 2](#_Toc195993182)

[4.2求解过程 2](#_Toc195993183)

[第5章 相关工作 3](#_Toc195993184)

# 第1章 题目概述

## 1.1 研究问题的背景

在数据库系统中，查询优化器的核心任务之一是估计查询的执行代价（如执行时间、资源消耗），以便选择最优的查询执行计划。传统的代价估计方法依赖于数据库的统计信息和启发式规则，但这些方法通常在复杂查询和大数据集场景下表现不佳。随着深度学习技术的发展，越来越多的学习型代价估计器被提出，这些方法通过训练数据学习到查询计划与代价之间的复杂关系，从而提供更高的预测精度。

本论文的主要贡献是提出了一种基于深度神经网络的查询代价估计方法，称为“Plan-Structured Deep Neural Network”（PS-DNN）。该方法通过将查询计划的结构信息与数据库元数据结合，利用深度学习模型预测查询的执行代价。

# 第2章 基础题的求解

## 2.1 求解思路

论文旨在通过深度学习模型提高数据库查询代价估计的精度，尤其是在复杂查询和大规模数据集上的表

现。

传统的代价估计方法在面对复杂查询时的表现不佳，论文提出的PS-DNN模型能够有效地处理查询计划中

的结构信息，提升代价预测的精度。

## 2.2 求解过程

论文提出的PS-DNN模型利用了查询计划的“计划结构”信息，将查询计划视为图结构，通过图神经网络

（GNN）来处理查询的树形结构。

该模型结合了数据库表的元数据（如表的大小、索引信息等），并将这些特征与查询计划一起输入到深

度神经网络中进行训练，最终输出查询的代价估计。

PS-DNN的核心是图神经网络（GNN），它能够处理查询计划中的树形结构，使得模型能够理解和利用

查询中各个操作节点之间的关系。

模型包括输入层（查询计划的特征）、多个隐藏层（通过深度学习提取特征）、以及输出层（预测查询

的代价）

数据集：论文使用了多个基准数据集进行实验，包括TPC-DS、IMDB JOB和STATS数据集。

实验设置：在实验中，PS-DNN模型与传统的基于统计信息的代价估计方法进行了比较，评估了模型在查

询代价预测精度、训练速度和推理速度上的表现。

精度：PS-DNN模型在各个数据集上的预测精度显著优于传统方法，尤其在复杂查询的代价估计上表现突

出。

训练速度：模型的训练速度与数据集的规模成正比，但整体训练过程相对较快，能够处理大规模数据。

推理速度：推理阶段的速度表现也较为优异，能够满足实时查询优化的需求。

# 第3章 高级题的求解

# 第4章 进阶题的求解

模型复杂性：尽管深度学习模型能够提供更高的精度，但其黑盒性质也带来了可解释性差的问题。在实

际应用中，数据库管理员可能需要更多的可解释性来理解模型的预测结果。

数据依赖性：模型的表现依赖于训练数据的质量和多样性，如果训练数据不能覆盖各种查询模式，模型

可能会在某些特定场景下出现偏差。

# 第5章 相关工作

这篇论文提出的PS-DNN模型在数据库查询代价估计中展示了深度学习技术的强大潜力，尤其是在处理复杂查

询和大规模数据集时，表现优于传统的统计方法。通过引入查询计划的结构信息和图神经网络，模型能够更好

地理解查询的执行过程，提供精确的代价估计。

然而，尽管PS-DNN在精度和推理速度上表现出色，但其黑盒特性和对数据的依赖性仍然是其主要局限性。未

来的研究可以在提升模型可解释性和扩展数据集多样性方面进行改进。

通过阅读这篇论文，我对深度学习在数据库领域的应用有了更深入的了解，也为我的项目提供了重要的参考。

未来我将考虑将PS-DNN模型的思想应用于我的工作中，尤其是在查询优化和代价估计方面。

**参考文献**

更多附录......