

## 论文题目

Efficient and Exact Query of Large Process Model Repositories in Cloud Workflow Systems

---

## 论文作者

H. Huang, R. Peng and Z. Feng

---

## 发表期刊信息

*IEEE Transactions on Services Computing*, vol. 11, no. 5, pp. 821-832, 1 Sept.-Oct. 2018.

---

## 技术问题

高效、准确地查询云 workflow 系统中的大型流程模型库。

---

## 现实背景

随着云计算平台被越来越多的企业和个人所广泛接受，底层云 workflow 系统积累了大量的业务流程模型。根据租户的需求检索和推荐最相似的流程模型变得非常重要，因为这不仅有利于促进现有模型资产的重用，而且有助于降低建模流程的错误率。由于云 workflow 存储库的规模越来越大，开发高效、准确的查询方法就显得尤为迫切。

---

## 作者思路

本文提出了一种改进的基于图结构的两阶段精确查询方法。在过滤阶段，采用任务的标签、join-attribute和split-attribute组成的复合任务索引来获取候选模型，大大减少了验证算法需要测试的流程模型数量。在验证阶段，提出了一种基于任务码的子图同构检验方法，对候选模型集进行了细化。结果表明，该方法能显著提高查询效率，缩短查询响应时间。

---

## 解决方案

- 为了提高任务索引的过滤能力，在以cnet表示的流程模型中，通过组合每个任务的join-attribute和split-attribute来构造一个复合任务索引。
  - 为了降低复杂子图同构算法在验证阶段的时间开销，利用任务的邻域信息和结构特征构造数据结构任务代码，提高子图同构算法的剪枝能力。
  - 在是否考虑任务相似性的基础上，提出了两种算法。
  - 在云 workflow 管理平台上评估方法，包括模型生成工具、模型导入工具、可视化建模工具、模型查询工具和结果显示工具。
- 

## 创新贡献

- 将任务的标签、join-attribute和split-attribute结合起来构建一个复合任务索引，提高过滤能力。
  - 提出了一种新的基于任务码的子图同构算法，该算法利用了任务节点的邻域信息和结构特征。
  - 对合成和真实数据集进行广泛的实验，以证明所提出方法的有效性和效率。
- 

## 效果评价

- 在不考虑任务相似性的情况下，QTCS可以极大地提高基于图结构的大型流程模型存储库中精确查询的效率。
- 无论是否考虑任务相似性，该方法都可以极大地提高基于大型流程模型存储库图结构的精确查询效率。

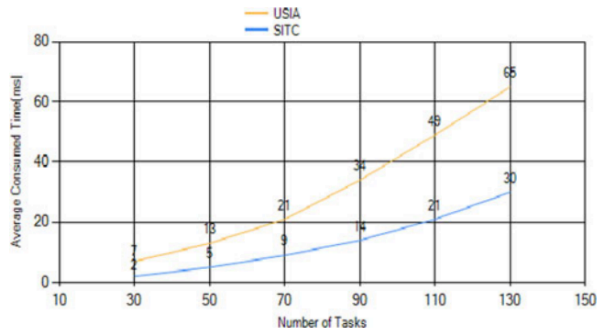


Fig. 5. Time comparison between SITC and USIA.

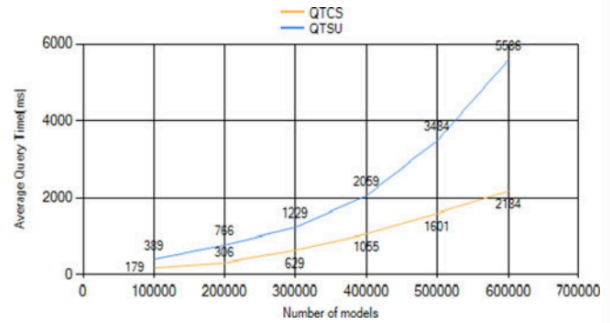


Fig. 6. Query time comparison without considering task similarity.

## 个人感想

### 本文的优势

本文针对云工作流系统中大型流程模型库的精确查询，提高了基于图结构的查询效率。本文的主要思想是受数据库管理技术领域综合索引概念的启发，利用减少候选模型的筛选阶段，提出了基于任务的子图同构测试的代码完善候选人模型和得到更精确的匹配模型。为了使方法更加灵活，本文考虑了任务相似性，以便在建模人员启动检索时为他们提供更多的选择。通过大量的实验对所提出的算法进行了评价。实验结果表明，该方法可以获得比现有方法更好的查询效率。

### 本文的劣势

- 为了进一步区分两个任务，尚未扩展复合任务索引，使其包含更多的任务属性，如数据或资源属性；
- 尚未在并行计算环境（如Hadoop生态系统）中扩展这个方法，以提高超大流程模型存储库的检索效率。