第5章 总结与展望

2. 1. 总结

随着业务流程管理技术的飞速发展， 越来越多的企业和组织引入各种各样的业务流程感知信息系统来支持其业务流程【引用】。流程挖掘技术就是伴随着业务流程管理技术的飞速发展而诞生并发展的。围绕流程挖掘，本文首先提出了一个可以挖掘复杂结构的流程挖掘算法,并针对流程挖掘算法评估构建了一组典型的模型集合。

流程挖掘技术肩负着业务流程管理中的发现、符合性检查和增强的作用。优秀的流程挖掘算法有效的支撑整个业务流程管理，然而现在的流程挖掘算法都存在着不能有效的挖掘出非自由选择结构与不可见任务嵌套的复杂结构。本文调研了当今的主流流程挖掘算法，对其优缺点进行了分析，同时重点分析了系列的流程挖掘算法，在此基础上设计并实现了可以挖掘不可见任务与非自由选择结构嵌套的流程挖掘算法。

流程挖掘算法整体思想是融合了系列中的和算法分别对非自由选择结构和不可见任务结构的处理，即算法在算法的基础上，分别挖掘事件日志中含有的虚假依赖和隐式依赖，达到能够挖掘含有不可见任务和非自由选择结构相嵌套的情况。然而简单的融合并不能正确的挖掘出正确流程模型，本文总结了在融合中出现的四个问题，并针对这四个问题给出了解决方案：发现改进的虚假依赖、补充可达关系、发现非自由选择结构和调整不可见任务。

本文介绍了在业务流程系统BeehiveZ中的实现，并且在手工数据集和真实数据集上分别对算法与主流流程挖掘算法进行对比实验。在实验中，我们采取了Fitness, Precision, Simplicity, Model rediscoverability等多个维度对比。实验表明可以正确的挖掘出含有非自由选择结构与不可见任务融合的流程模型，并且在各个维度上挖掘性能较主流流程挖掘算法较好。

伴随着流程挖掘技术的发展，针对流程挖掘的评估也成为当下研究的热点。流程挖掘的评估框架可以为帮助企业选择出最合适的流程挖掘算法，现阶段已经有基于专有模型全集和基于专有模型子集与推荐技术的流程挖掘框架。然而这些流程挖掘框架面临着评估时间长，计算开销大，评估结果受限于流程专有子集的质量。因此本文针对上述情况，构建了一个典型模型库，优化流程挖掘评估框架。

本文对描述流程模型的特征进行降维，将48维特征通过两条描述流程模型的标准降至6维，即不可见任务个数、重名任务个数、非自由选择结构个数、非结构化循环个数、短循环个数和嵌套循环个数。降维使得评估框架中对流程模型的刻画表现的更加准确。同时，依据这六条属性，本文构建了针对流程挖掘评估的典型模型库。针对每个属性，模型库中均手工构造10个包含对应特征的流程模型，总计60个流程模型。分别针对模型降维和典型模型库在手工和真实流程模型数据中进行实验，实验表明本文提出的改进可以在改进流程挖掘算法评估框架中评估时间长、计算量大的问题的同时流程挖掘评估的准确度。

* 1. 展望

围绕流程挖掘，本文的工作为一个流程挖掘算法和构建典型模型库，对于未来工作，有如下几点考虑：

1. 对本文提出的流程挖掘算法的正确性、挖掘能力和不足进行理论分析。即从理论层面推理本文的算法，分析算法的能力，分析算法在哪些事件日志中可以正确的挖掘出流程模型，找出算法的不足，并尽可能解决挖掘算法的局限性。
2. 设计可以并发操作的流程挖掘算法。针对一个事件日志，算法只能在单机下挖掘。算法的挖掘效率受限于一台单机的计算能力。尝试将算法中的算法流程做分解，尽量可以并行的挖掘。
3. 流程挖掘算法对事件日志的要求是完整且不含噪音，接下来的工作可以从对事件日志的预处理方向考虑。如设计算法，对不完整或者含有噪音的事件日志进行处理，使得其过滤修复为完整且不含有噪音。
4. 发现更多的影响流程挖掘的流程模型属性并更新典型模型库。本文我们找到6维描述流程挖掘的模型属性。进一步深入分析现今主流挖掘算法，找出更多的可以区分不同流程挖掘算法的模型属性，并且依据这些属性对典型模型库更新。