

TiD2016
质量竞争力大会
软件研发顶级盛会



过程挖掘

高燕

赛仕软件北京研发中心

yan.gao@sas.com

提纲

- 过程挖掘的必要性
- 过程挖掘的定义
- 过程挖掘的实现
- 过程挖掘的应用

为什么要挖掘过程？

企业不能满足客户的期望 **85%** 归因于系统和**流程**的缺陷，而非员工。

-- 质量管理大师戴明博士

过程挖掘的目的



成功案例(1/2)

- 荷兰 WoonFriesland 房屋租赁公司



Savings

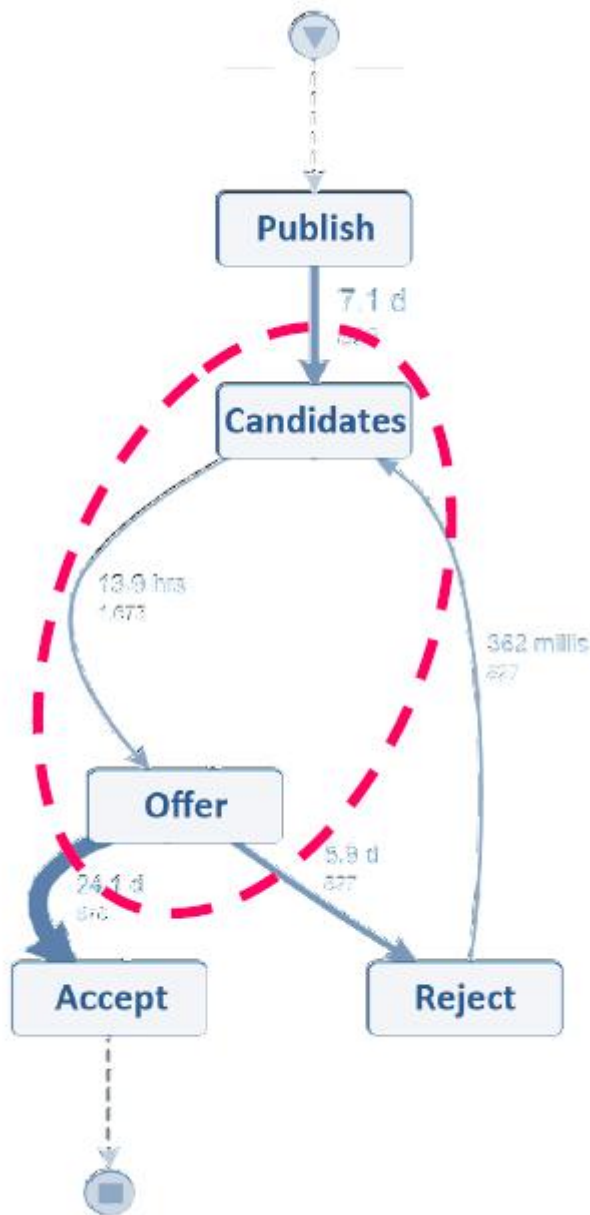
600 cases x 7,5 days = 4500 days faster
 4500 days x € 15 = € 67.500

Total costs

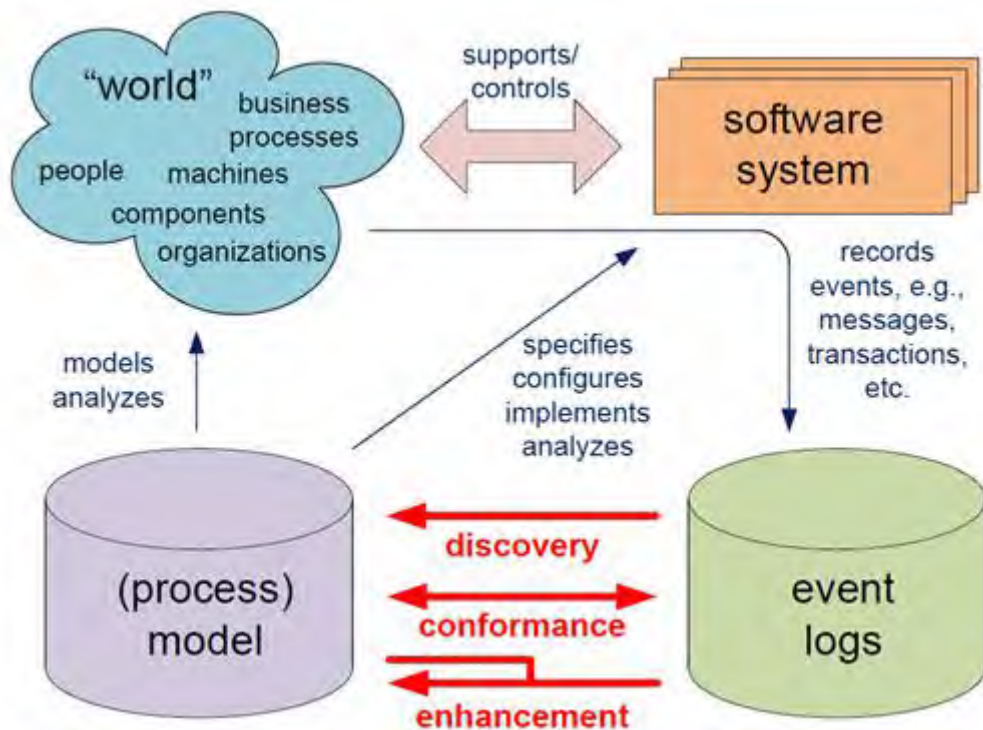
600 cases x 17 days x € 15 = € 153.000

成功案例(2/2)

- 过程瓶颈分析
 - 等候员工处理
 - 等待人工报价
 - 等待租房者线下答复
- 过程改进措施
 - 等候邮件通知
 - 自动报价
 - 在线接受/拒绝

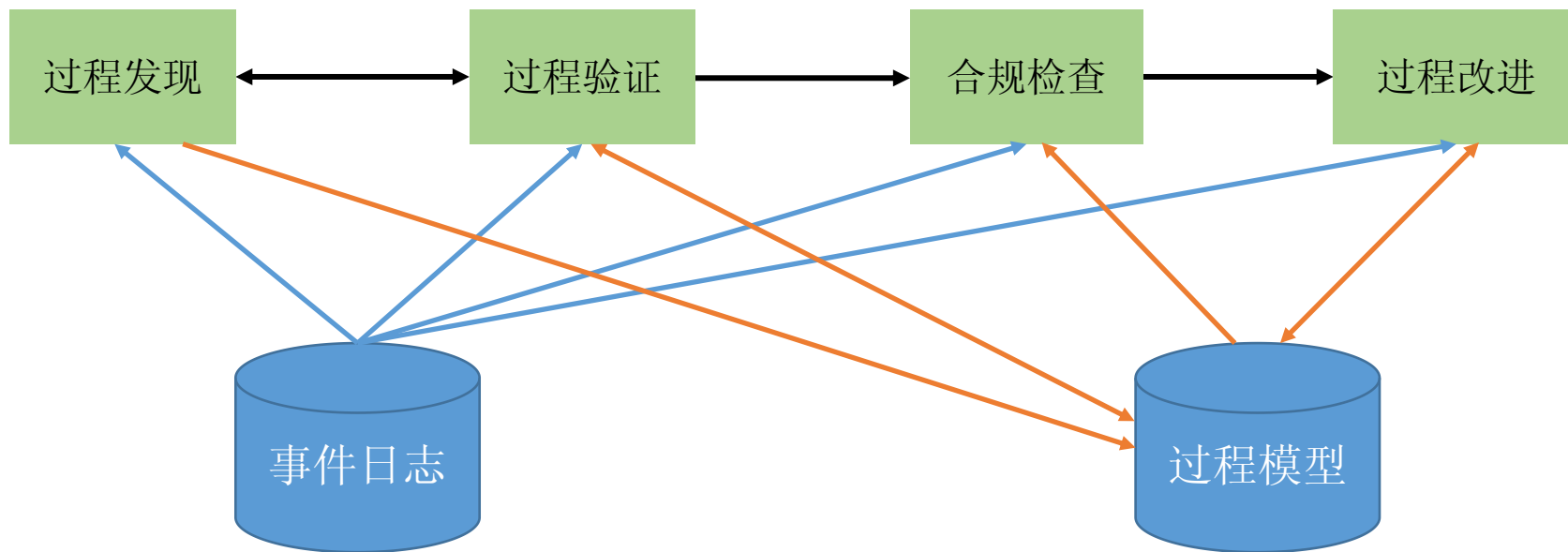


过程挖掘



- 过程发现
 - 理解实际的过程
- 合规检查
 - 检测不合规操作
 - 测量不合规程度
- 过程改进
 - 识别瓶颈
 - 识别零增值环节

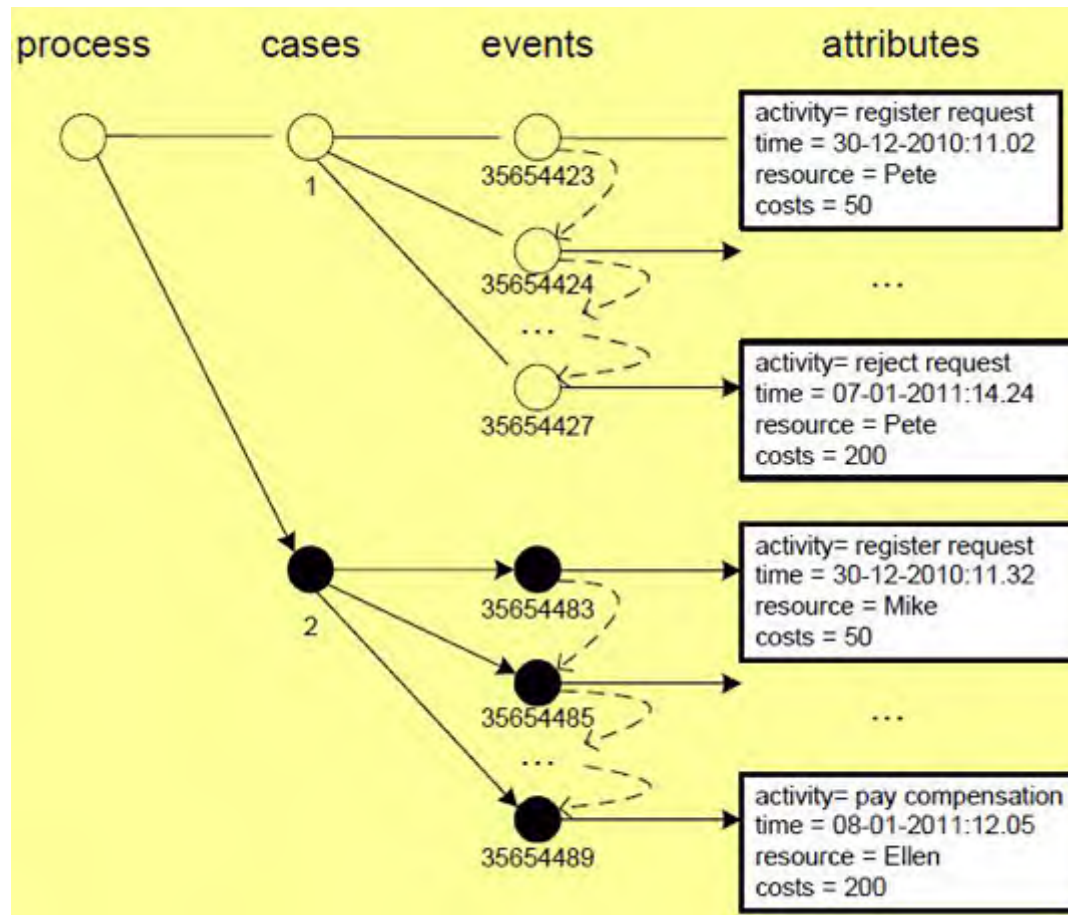
分析流程



事件日志

- 记录业务操作的基本信息
 - 案例标识号、事件标识号、事件描述、发生时间
- 其他对过程挖掘有用的信息
 - 事件执行人、执行成本等信息

事件日志示例



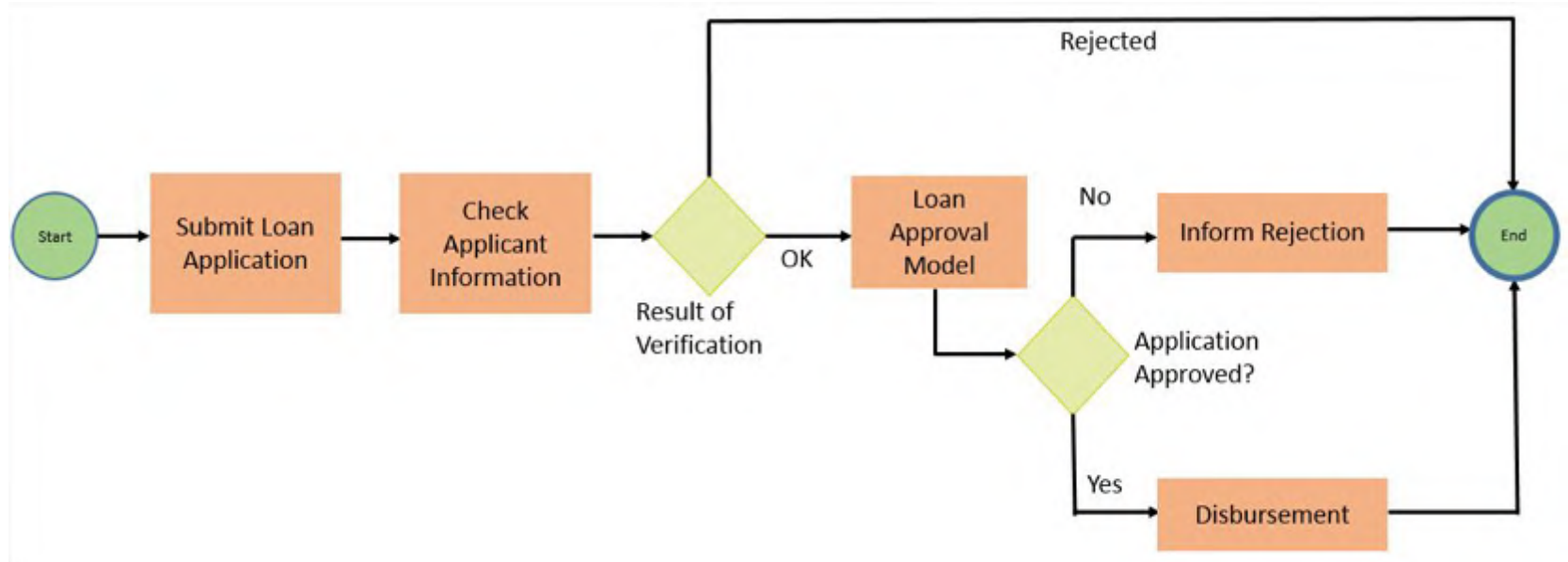
数据存储和交换标准

- MXML (Mining eXtensible Markup Language)
 - 出现于 2003 年
 - 难以扩展
- XES (eXtensible Event Stream)
 - 2010 年 9 月被 IEEE Task Force on Process Mining 小组采用。
 - 可扩展性更好

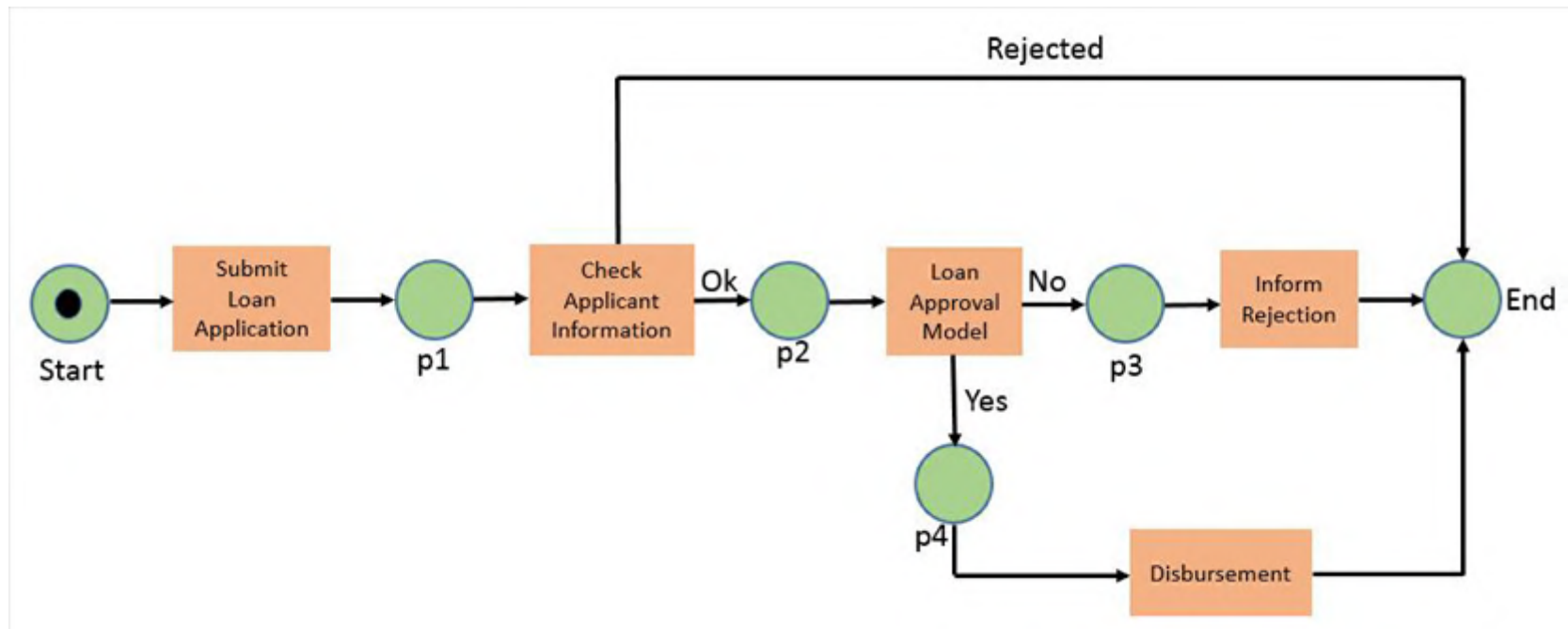
XES 标准示例

```
<trace>
  <date key="REG_DATE" value="2011-10-01T16:39:55.798+02:00"/>
  <string key="concept:name" value="173772"/>
  <string key="AMOUNT_REQ" value="3000"/>
  <event>
    <string key="org:resource" value="112"/>
    <string key="lifecycle:transition" value="COMPLETE"/>
    <string key="concept:name" value="A_SUBMITTED"/>
    <date key="time:timestamp" value="2011-10-01T16:39:55.798+02:00"/>
  </event>
  <event>
    <string key="org:resource" value="112"/>
    <string key="lifecycle:transition" value="COMPLETE"/>
    <string key="concept:name" value="A_PARTILYSUBMITTED"/>
    <date key="time:timestamp" value="2011-10-01T16:39:55.898+02:00"/>
  </event>
  <event>
    <string key="org:resource" value="112"/>
    <string key="lifecycle:transition" value="COMPLETE"/>
    <string key="concept:name" value="A_DECLINED"/>
    <date key="time:timestamp" value="2011-10-01T16:40:31.075+02:00"/>
  </event>
</trace>
```

模型表示法 - BPMN



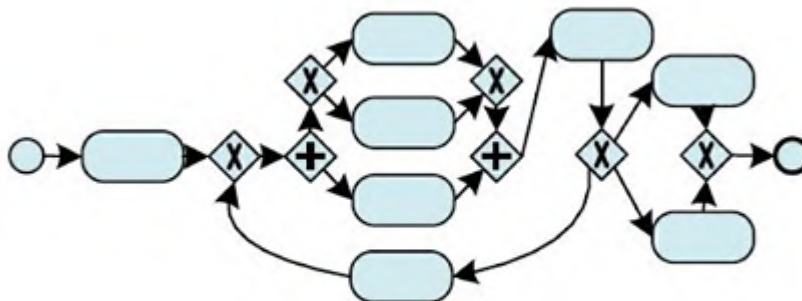
模型表示法 – Petri Nets



过程发现



event log



process model

过程发现算法

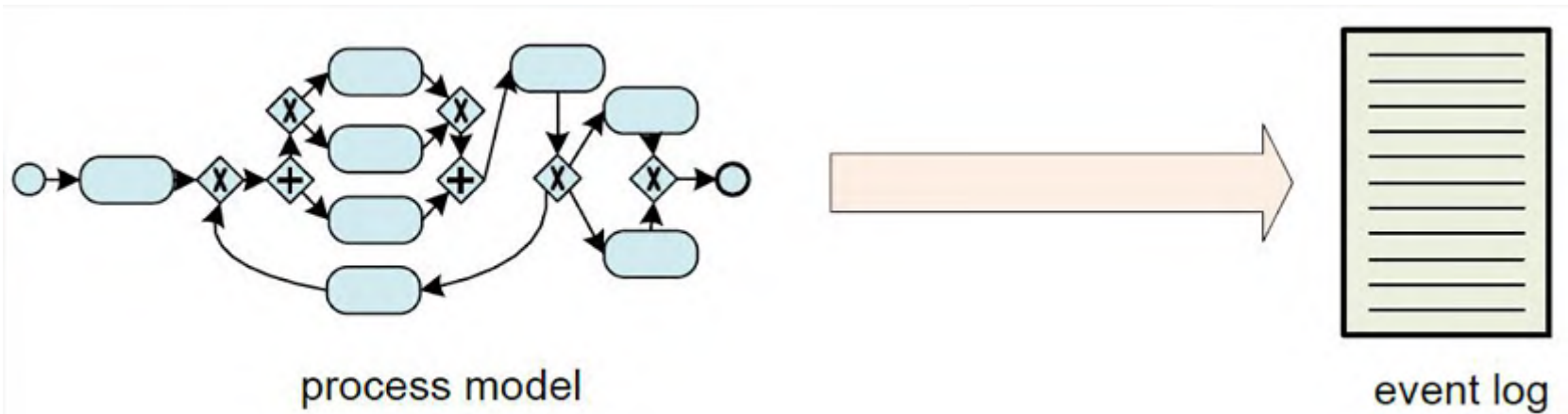
- 标识日志文件
- 过滤未完成案例和无用的事件
- 将业务活动及其所处的生命周期信息合并为事件
- 对事件日志按照事件发生的时间戳排序，记录每个案例中事件发生的顺序。
- 借助数据挖掘软件的顺序发现算法发现过程模型
- 人工验证、纠正和改进发现的过程模型
- 评估过程模型拟合的程度

核心算法实现

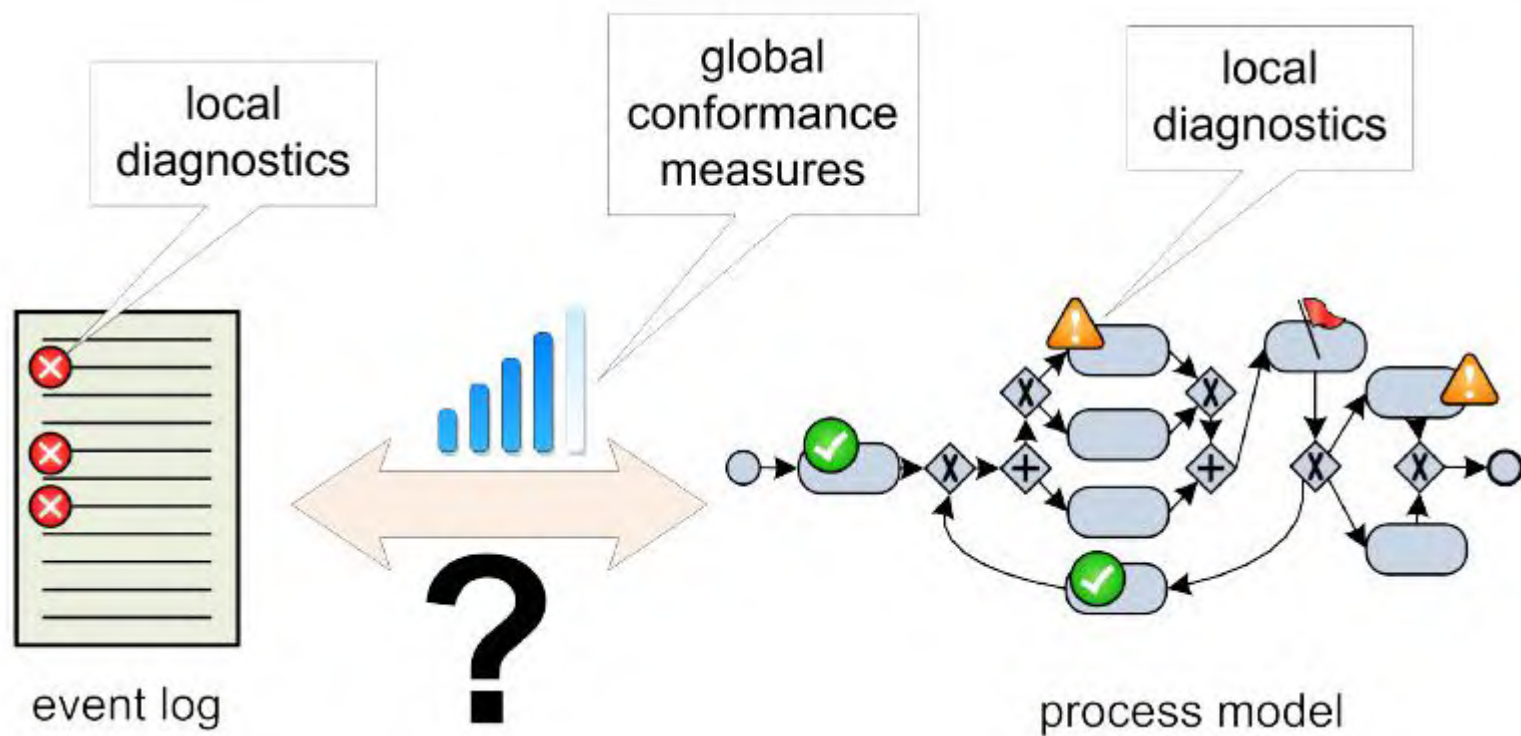
- 借助数据挖掘软件发现显著的业务过程路径
- 将过程路径转换成业务流模型

```
proc dmdb data=event_logs dmdbcat=catSeq;  
    id case_id;  
    class event;  
run;  
  
proc assoc data=event_logs dmdbcat=catSeq items=2 out=assocOut support=0;  
    customer case_id;  
    target event;  
run;  
  
proc sequence data=event_logs dmdbcat=catSeq assoc=assocOut out=seqOut  
nitems=2 support=0;  
    customer case_id;  
    target event;  
    visit visit/window=1;  
run;
```

过程执行



合规检查



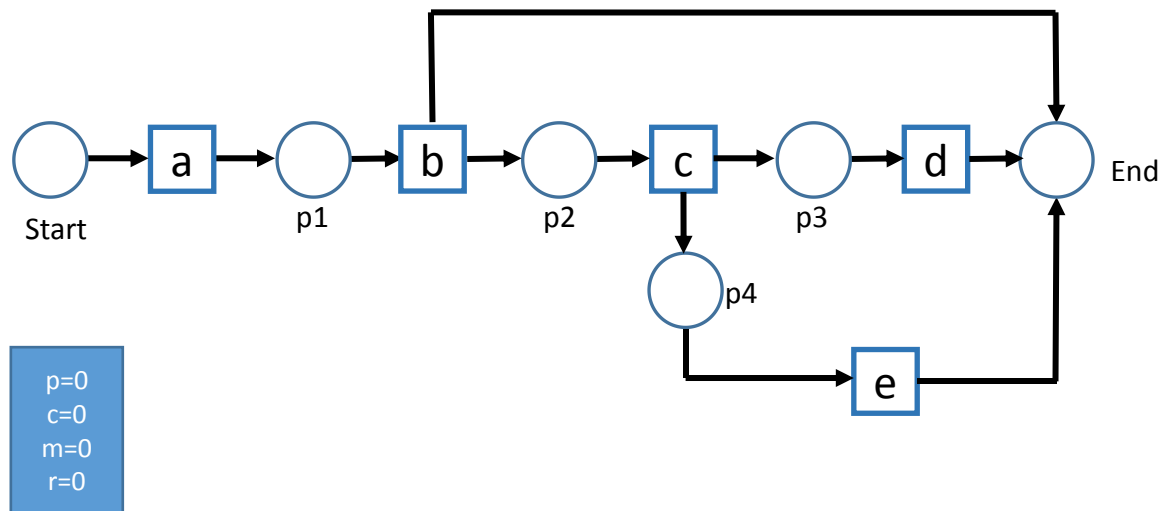
合规检查算法

- 将过程模型转换成 Petri Nets 格式
- 基于 Petri Nets 重演事件日志，并为每个案例记录下面四个测度。
 - 缺失的托肯数 (m)
 - 消耗的托肯数 (c)
 - 遗留的托肯数 (r)
 - 产生的托肯数 (p)
- 将所有案例数据的托肯数按照以上四种类型分别汇总
- 计算业务过程模型总体的拟合度

$$\text{拟合度} = \frac{1}{2} * (1 - m/c) + \frac{1}{2} * (1 - r/p)$$

合规检查算法示例

待重演的事件日志: {a, b, d}

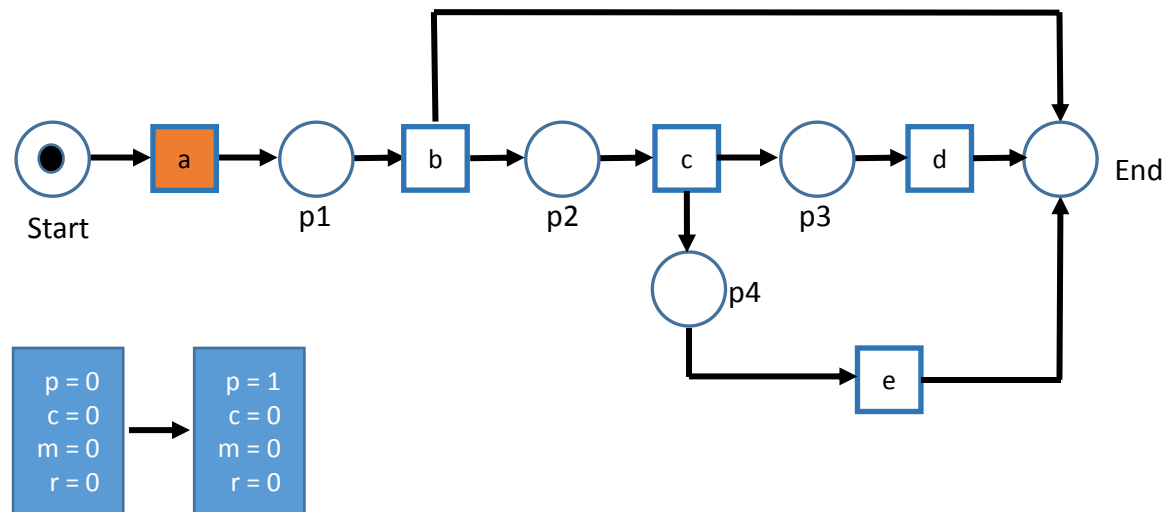


产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

合规检查算法示例

待重演的事件日志: {a, b, d}

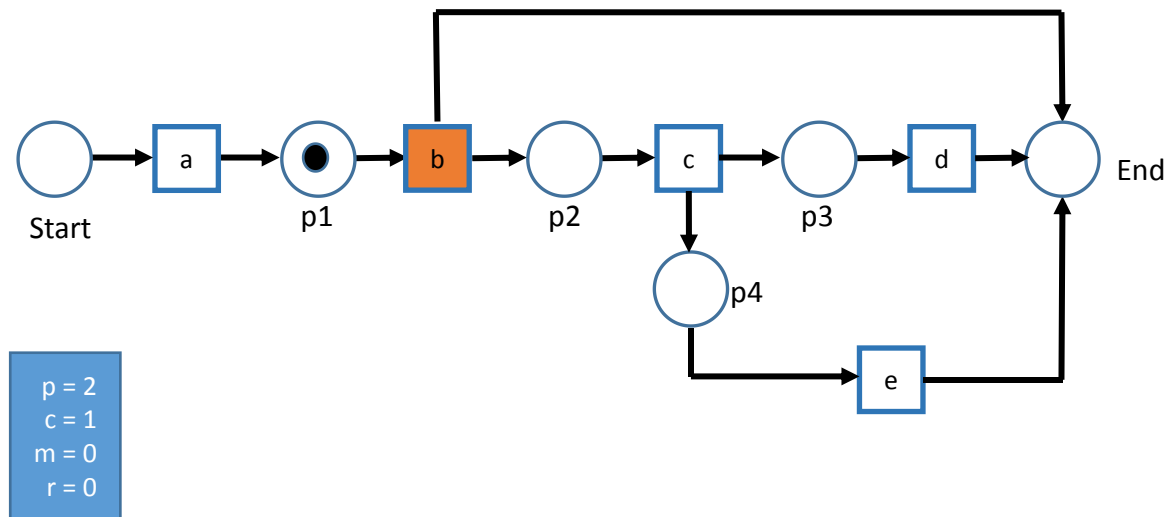


产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

合规检查算法示例

待重演的事件日志: {a, b, d}

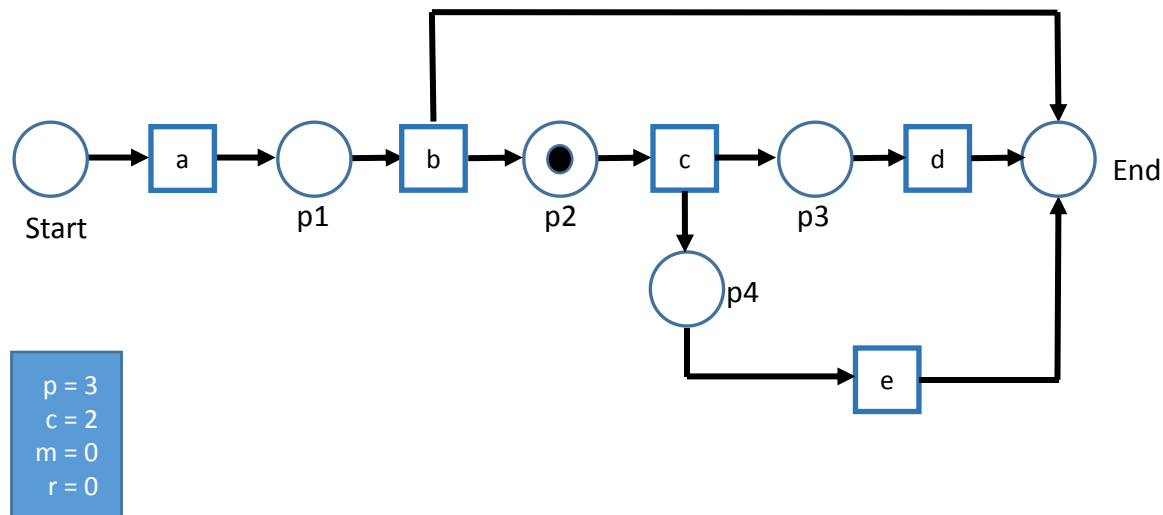


产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

合规检查算法示例

待重演的事件日志: {a, b, d}

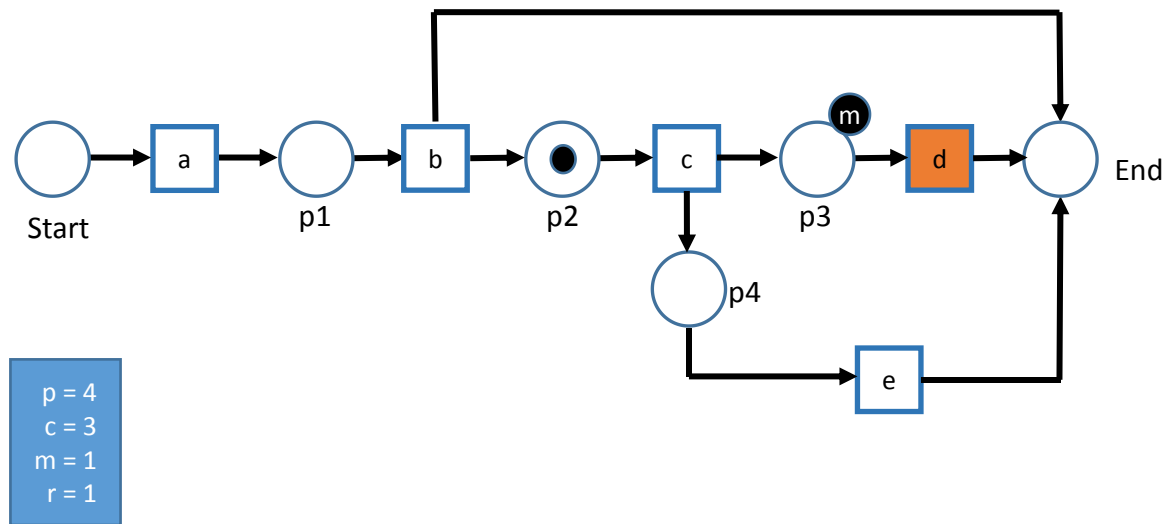


产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

合规检查算法示例

待重演的事件日志: {a, b, d}

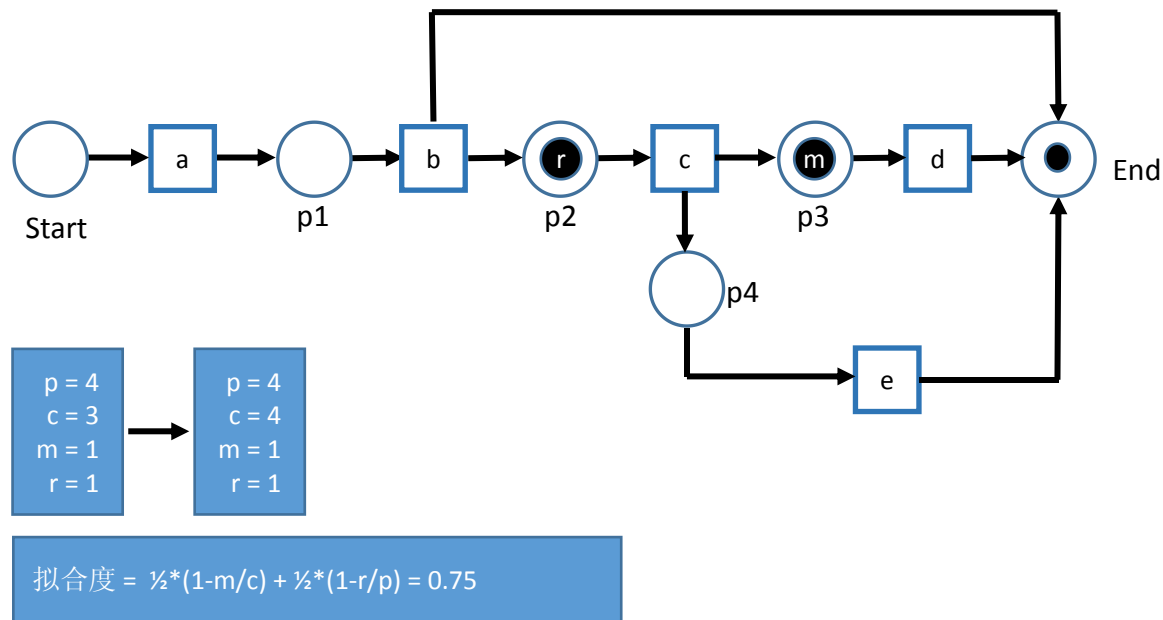


产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

合规检查算法示例

待重演的事件日志: {a, b, d}



产生的托肯数 (p)
缺失的托肯数 (m)

消耗的托肯数 (c)
遗留的托肯数 (r)

案例分享

- 背景
 - 每年上百个项目
- 需求
 - 风险项目标识
 - 问题识别和原因剖析

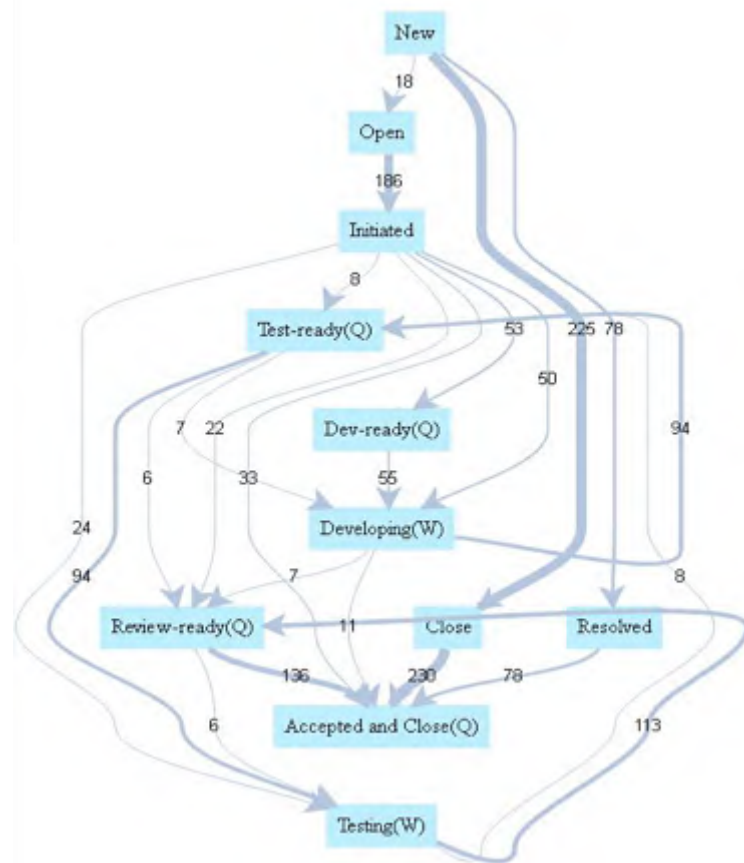
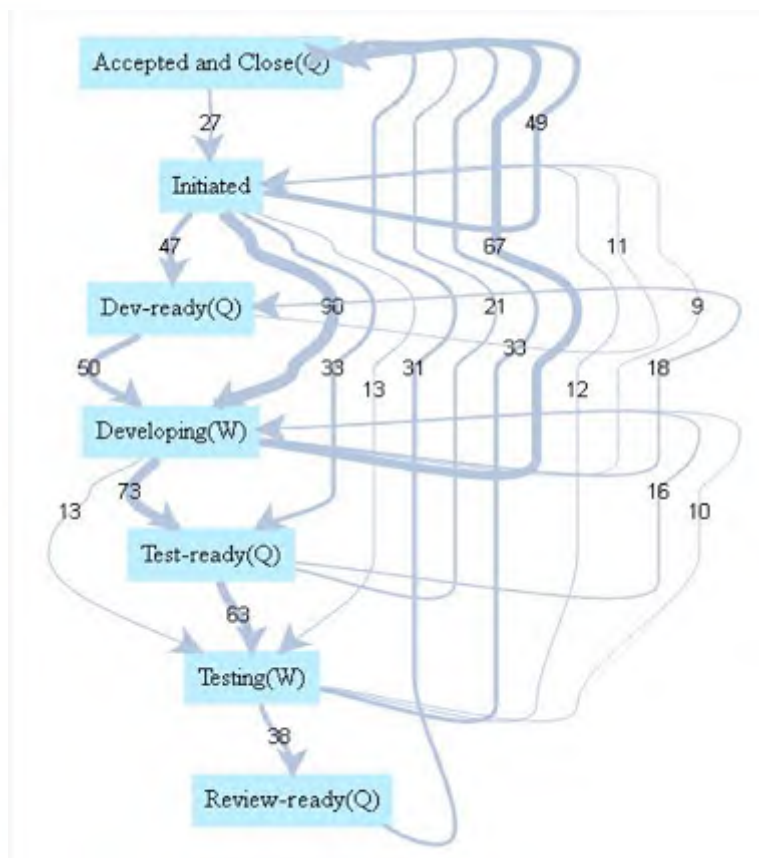
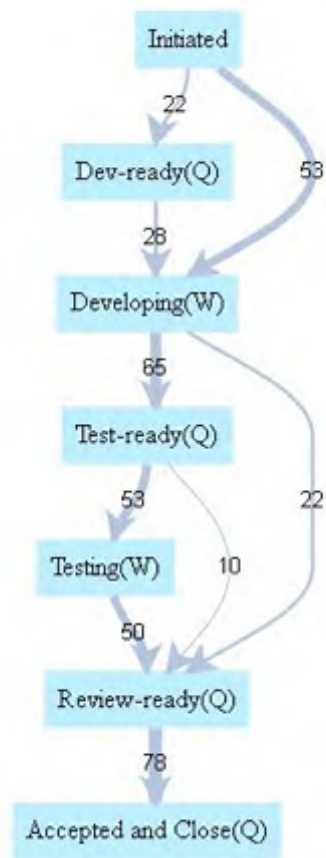
数据

- JIRA 项目数据
- 30 个项目
- Epic
 - Epic ID
 - Epic Name
- Story
 - Story ID
 - Story Name
 - Owner
 - Status
 - Priority
 - Epic ID
 - Fix Versions
 - Created
 - Updated
- Story history
 - Story ID
 - Field
 - From String
 - To String
 - Display Name
 - Updated
 - Sequence

过程发现和验证(1/2)

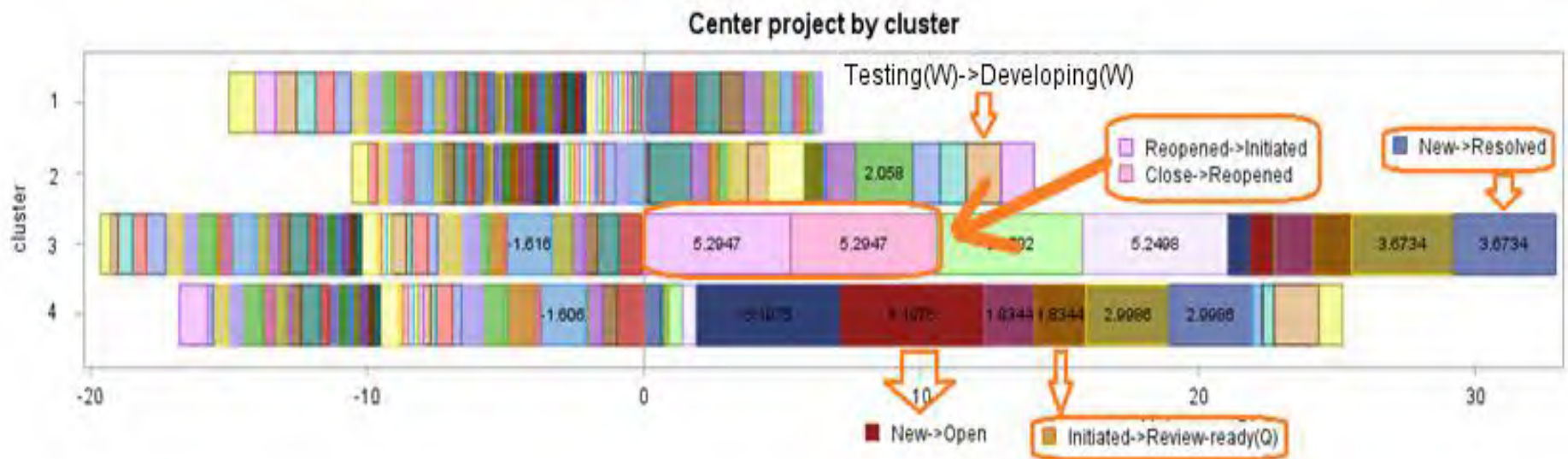
- 标准过程定义过于灵活
- 实际过程多种多样
- 过程执行中的两大问题
 - 走捷径
 - 后退

过程发现和验证(2/2)



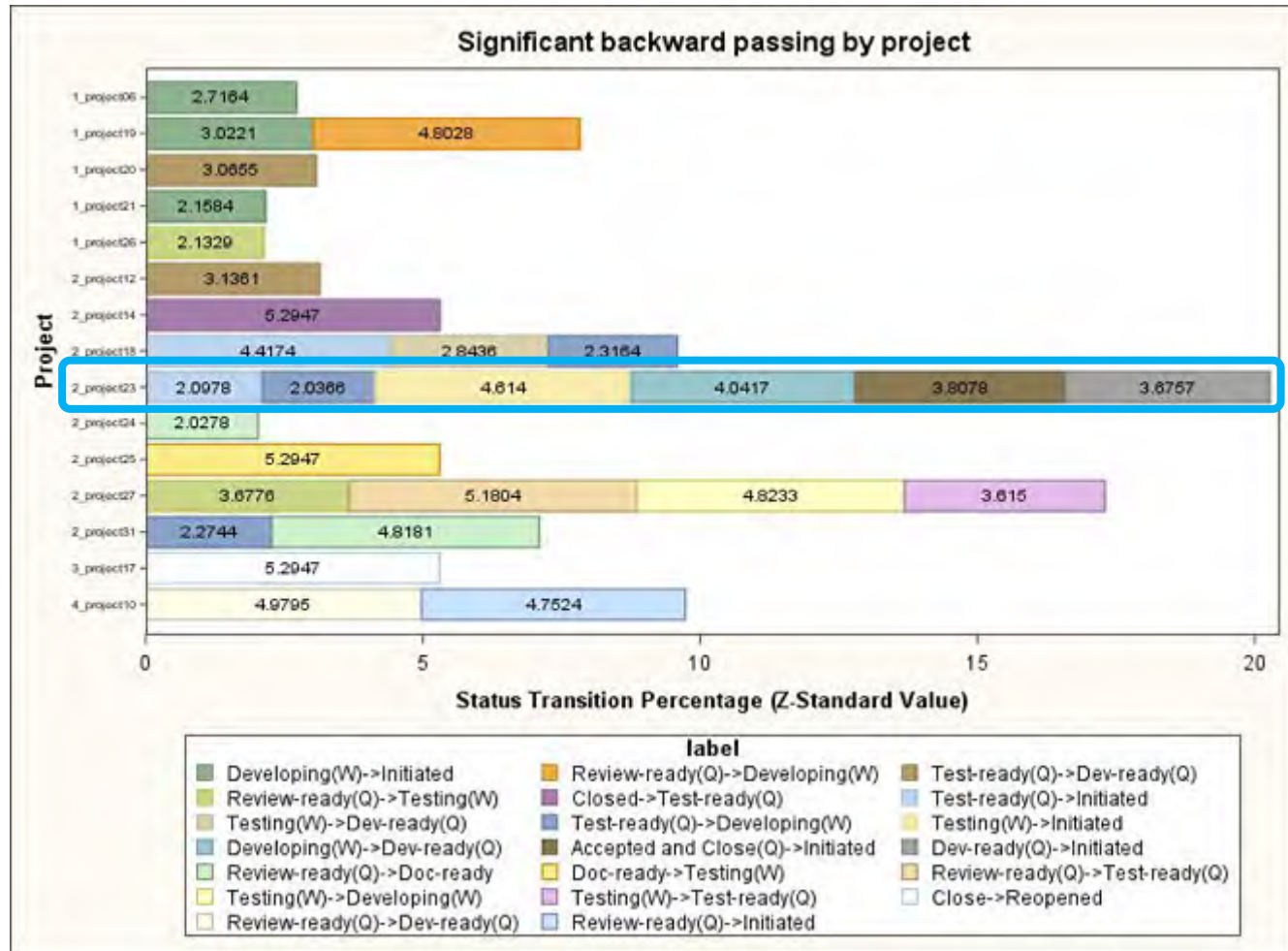
项目聚类分析

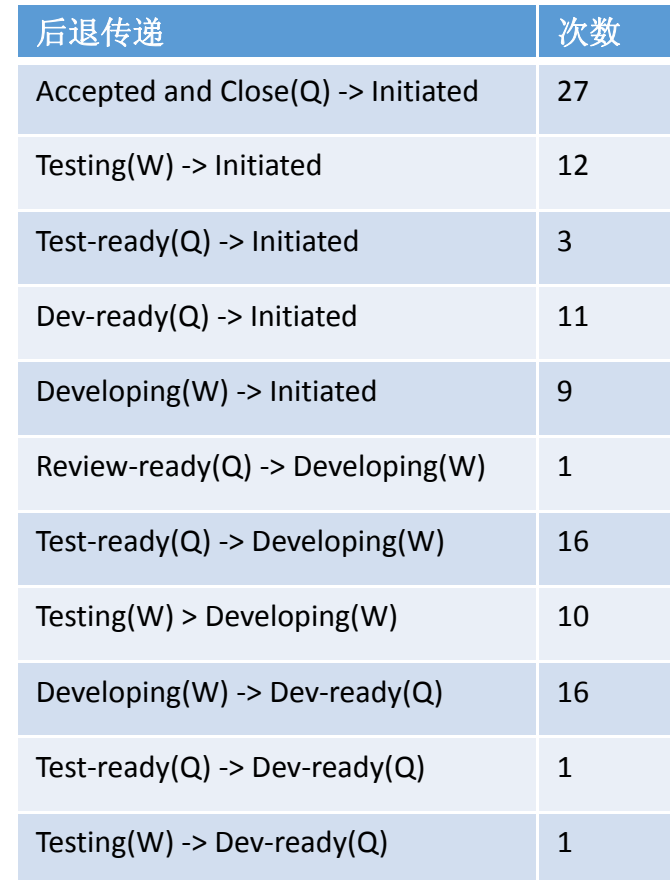
- 第一类：遵循标准过程，过程基本正常
- 第二类：遵循标准过程，过程频繁后退
- 第三类、第四类：非标准过程



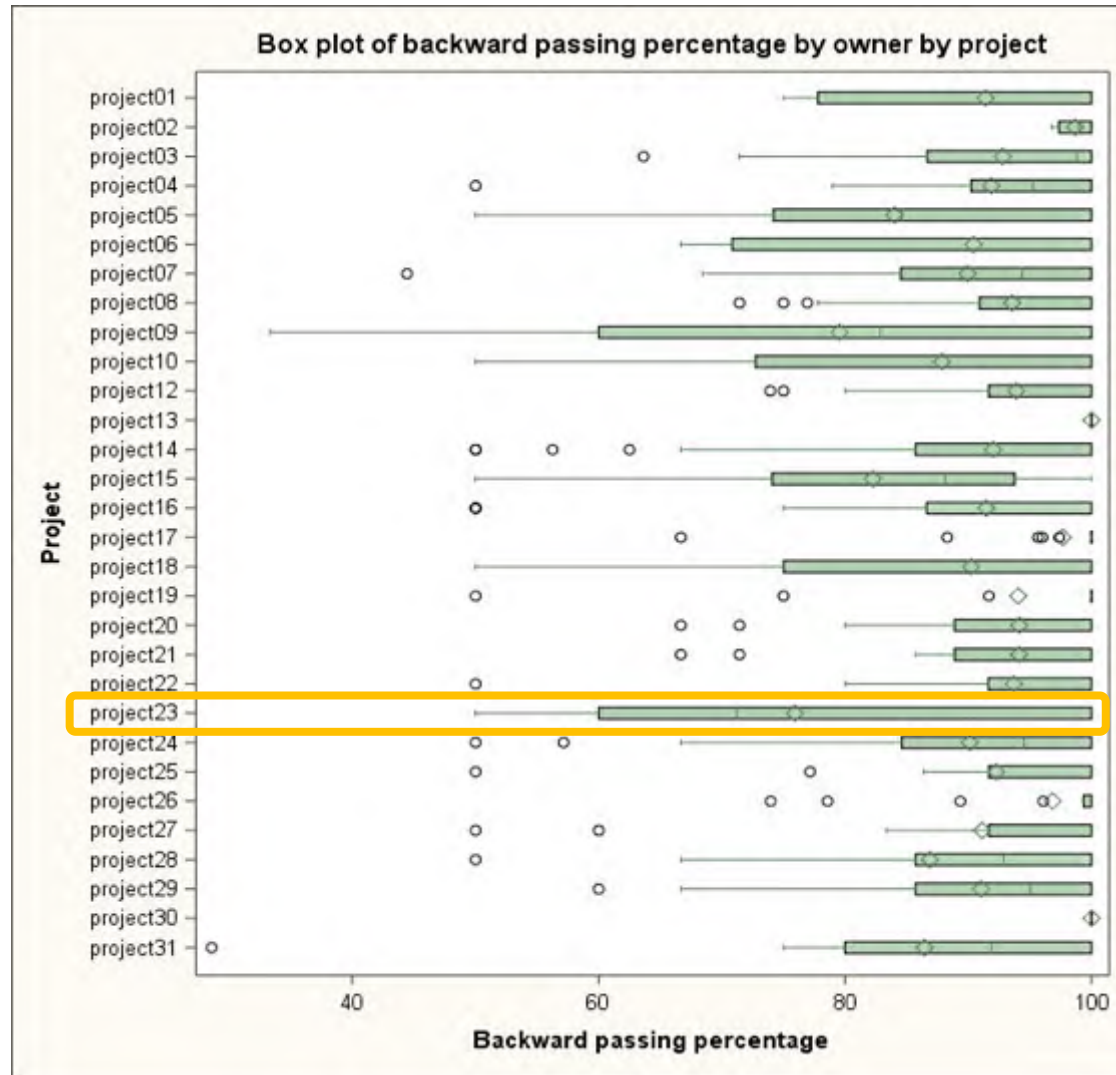
风险项目标识

- 过程频繁后退的项目





问题识别 (2/2)



总结

- 过程挖掘的必要性
- 过程挖掘的定义
- 过程挖掘的实现
- 过程挖掘的应用

高燕
赛仕软件北京研发中心
yan.gao@sas.com