































进行分解等.这样,在重用传统工作流时,可以通过检索行为相似的 stream 来修正检索工作流.

### 3 实验与结果分析

#### 3.1 实验设计

本文的实验基于如下环境: Intel Core(TM) CPU2.2 GHz, 8.0GB RAM, Windows7 64 位操作系统, JDK1.6 环境的计算机. 目前, 专用的语义工作流的实验数据集还较少. 本文的实验数据集选自 WikiTaaable(<http://wikitaaable.loria.fr>)的 Recipe 和 Recipesource(<http://www.recipesource.com>), 其中, 共有 1 000 多个意大利面食 (pasta) 食谱. 实验中的领域任务本体 *TaskOnto* 来自 WikiTaaable 的 Culinary actions 本体, 有 100 个左右节点; 领域数据对象本体 *DataOnto* 来自 WikiTaaable 的 Food 本体, 有 200 个左右节点. WikiTaaable, Recipesource 及其中的本体是 ICCBR 会议为计算机烹饪比赛 (computer cooking contest, 简称 CCC) 建立的.

本文从意大利面食食谱数据集中选取 100 个食谱<sup>[14]</sup>, 根据定义 1, 将食谱的烹饪说明 (cooking instruction) 表示成如图 1 的语义工作流<sup>[1,14]</sup>, 组成规模为 100 的烹饪语义工作流库 *SWB<sub>0</sub>*. 本实验从 *SWB<sub>0</sub>* 中随机选取 50 个语义工作流, 组成测试语义工作流库 *SWB<sub>1</sub>*. *SWB<sub>1</sub>* 的基本结构特征见表 1.

**Table 1** Basic characteristics of the case base of culinary semantic workflows

表 1 测试烹饪语义工作流库的基本特征

数据集	平均任务节点数	平均数据节点数	平均路由节点数	平均控制边数	平均并发分支数
烹饪语义工作流库	7.8	12.3	3.2	14.6	2.5

同时, 请 3 位烹饪专家组成领域专家组来评价检索到的匹配工作流 stream 和修正语义工作流的质量.

#### 3.2 本文的修正方法与基于结构特征的修正方法比较

首先开展实验 1: 比较基于行为特征的 stream 相似性方法和基于结构特征的 stream 相似性方法<sup>[2]</sup>对相似的匹配 stream 的检索性能. 从 *SWB<sub>1</sub>* 任选 10 个工作流组成测试工作流集 *QSW<sub>1</sub>*. 对每一个测试语义工作流, 任选它的一个 stream 记录下来. 对该 stream 做如下改变: 如果为顺序结构, 则为它的某个任务节点添加包含一个语义描述相似的任务节点的并发和选择分支; 如果为并发或选择结构, 则任选其中的两个任务节点, 将其改变为语义描述相似的任务节点. 这样共记录 10 个 stream 并组成检索 stream 集 *QST*, 同时可以得到 20 个改变后的语义工作流. 这 20 个改变后的语义工作流与余下的 40 个语义工作流一起组成规模为 60 的新语义工作流库 *SWB<sub>2</sub>*. 对 *SWB<sub>2</sub>* 中的每个语义工作流进行分解, 共得到 215 个 stream, 组成工作流 stream 库 *STB<sub>1</sub>*.

针对检索 stream 集 *QST*, 在 *STB<sub>1</sub>* 中检索相似的匹配 stream. 具体如下: 对于 *QST* 中的每个 stream (称为查询 stream), 首先, 在 *STB<sub>1</sub>* 中检索匹配 stream 集合 *MS<sub>1</sub>*; 然后, 在 *MS<sub>1</sub>* 中检索行为特征相似的 stream.

对于定义 12 的 stream 的行为相似性, 采用基于相似性的检索 (similarity based retrieval) 方法执行上述检索过程. 记录每个查询 stream 的检索结果集中的相似匹配 stream 的数量, 并对所有查询 stream 的相似匹配 stream 的数量求均值. 使用基于结构特征的修正方法<sup>[2]</sup>中的 stream 相似性方法执行相同的检索过程. 检索结果见表 2.

**Table 2** Comparisons of retrieval results of similar matching workflow streams

表 2 相似的匹配工作流 stream 的检索结果比较

算法	基于结构特征的相似性	基于行为特征的相似性
相似的匹配工作流 stream 的平均数量	2.1	3.3

由表 2 得, 基于行为特征的相似方法的检索结果集中, 匹配 stream 的平均数量高于基于结构特征的相似方法<sup>[2]</sup>. 这是因为基于 stream 行为特征的方法不但检索出了结构相似的匹配 stream, 而且检索出了行为相似的匹配 stream. 这样可以在结构相似的匹配 stream 不存在的情况下, 选用行为相似的匹配 stream 来完成语义工作流修正任务.

同时, 请领域专家组评价检索到的 stream 的质量, 分别绘制两种相似性方法对应的 *P-R* 曲线. 如图 13 所示.

*P-R*(precision-recall)曲线是评价检索系统的整体检索性能的方法之一.整体检索性能较好的检索系统的 *P-R* 曲线全部或绝大部分处于整体检索性能一般的检索系统的 *P-R* 曲线上方.

由图 13 可以得出,基于行为特征相似性的匹配 stream 检索方法的 *P-R* 曲线整体处于基于结构特征相似性的检索方法<sup>[2]</sup>的 *P-R* 曲线上方.这说明基于行为特征相似性的匹配 stream 检索方法的检索性能要好于基于结构特征相似性的检索方法<sup>[2]</sup>.

接下来开展实验 2:检验本文的基于 stream 行为特征的工作流修正方法是否执行了变更要求.从测试工作流集  $SWB_2$  中任选 10 个工作流组成测试工作流集  $QSW_2$ .针对  $QSW_2$  中的每个测试工作流  $TW_i$ ,首先删除 1~2 个数据对象节点,并将其 1~2 个数据对象节点的语义描述替换为相似的语义描述,得到语义工作流  $TW'_i$ .在  $SWB_2$  中,用  $TW'_i$  替换  $TW_i$ ,然后,对  $QSW_2$  中的每个测试工作流  $TW_i$ ,使用语义工作流的结构相似性方法<sup>[1]</sup>在  $SWB_2$  中检索一个相似的语义工作流  $RW_i$ ,以使得可以通过执行一组删除(delete)和加入(add)数据对象节点的操作将  $RW_i$  转变为  $TW_i$ <sup>[2]</sup>.这些删除和加入操作由手工得出,作为工作流修正时的变更要求 *chReq*.然后,对于每个语义工作流  $RW_i$  及其对应的 *chReq*,执行本文的工作流修正算法,可以得到 10 个修正语义工作流.分析 10 个修正语义工作流,并请领域专家组对它们进行评价.结果表明,测试语义工作流均完成了修正过程,并且与变更请求 *chReq* 基本一致.

最后开展实验 3:比较基于 stream 行为特征的修正方法和基于 stream 结构特征的修正方法<sup>[2]</sup>得到的修正工作流的质量.将实验 2 中的测试工作流集  $QSW_2$  中的每个测试工作流  $TW_i$  与其对应的修正语义工作流  $RW_i$  混在一起,这样得到包含 20 个工作流的工作流集合  $RWS$ .在事先不告诉每个工作流是测试工作流还是修正语义工作流的前提下,请 3 位领域专家使用 Likert 评分方法(1~5 分)对  $RWS$  中的每个工作流的合理性进行评分.然后将每个测试工作流  $TW_i$  和对应修正语义工作流  $RW_i$  组成一对( $TW_i, RW_i$ ),这样共得到 30 对.比较每对( $TW_i, RW_i$ )中两个工作流评分的高低.同时,对测试工作流集  $QSW_2$  中的每个语义工作流  $TW_i$  及其对应的变更要求 *chReq*,使用基于 stream 结构特征的修正方法<sup>[2]</sup>执行工作流修正操作.对得到的修正工作流进行相同的评分过程.结果如图 14 所示.

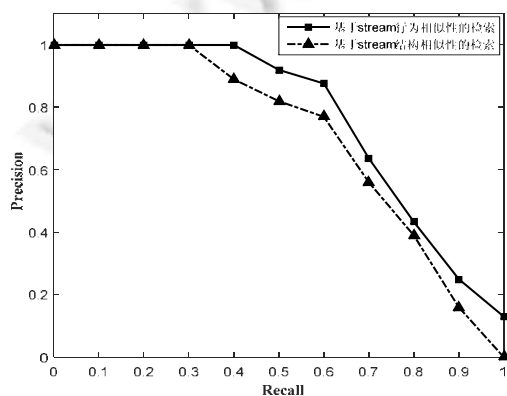


Fig.13 Comparisons of retrieval performances on workflow streams

图 13 工作流 stream 的检索性能比较

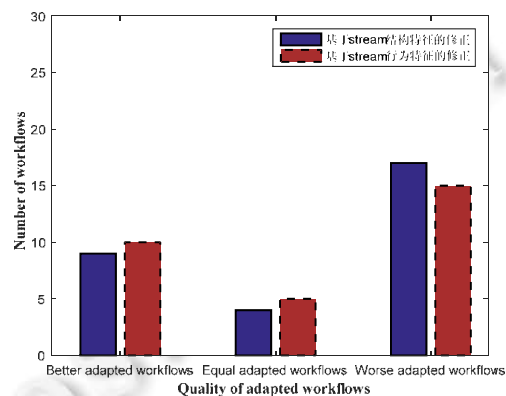


Fig.14 Comparisons of qualities of adapted semantic workflow

图 14 修正语义工作流的质量比较

图 14 的横坐标分别为在语义工作流修正结束后,领域专家对修正语义工作流的评分高于、低于对测试语义工作流的评分和领域专家对二者的评分相等这 3 种情况;纵坐标为 3 种情况对应的修正工作流数量.

由图 14 可以看出,对于基于 stream 行为特征的修正方法,共有 15 个修正语义工作流的评分不低于(高于或等于)测试工作流,比例为 50%.而基于 stream 结构特征的修正方法<sup>[2]</sup>共有 13 个修正语义工作流的评分不低于(高于或等于)测试工作流,比例约为 43%.基于行为特征的修正方法的评分比基于结构特征的修正方法高出 7%.从而与基于 stream 结构特征的修正方法<sup>[2]</sup>相比,基于 stream 行为特征的修正方法得到了整体质量更好的修正语



义 workflow 集合。

从以上 3 个实验可得:与基于 stream 结构特征的修正方法<sup>[2]</sup>相比,基于 stream 行为特征的修正方法提高了修正算法的适应能力,较好地改善了修正语义 workflow 的质量.同时也可以得出:基于 stream 结构特征的修正方法<sup>[2]</sup>更适用于科学 workflow 的修正,而本文的基于 stream 行为特征的修正方法更适用于业务 workflow 的修正。

#### 4 总结与展望

本文提出了一种不需要修正知识、基于 workflow stream 行为特征的语义 workflow 修正算法,该算法适用的前提是具有一定规模的语义 workflow 库.该算法以引入领域知识的块结构化语义 workflow 为研究对象,使用任务紧邻关系集刻画 workflow stream 的执行行为.结合领域数据知识构造了 workflow stream 库的锚集合数据索引 ASDataIndex.针对检索语义 workflow 中的每个查询 workflow stream,先基于 ASDataIndex 和 workflow stream 匹配规则对 workflow stream 库进行过滤,得到候选匹配 workflow stream 集;然后,基于行为特征的相似性和变更请求对候选 workflow stream 集进行验证,得到与当前变更请求一致程度最高和相似的匹配 workflow stream;接着,使用检索到的匹配 workflow stream 替换原查询 workflow stream,逐步对检索语义 workflow 进行修正;最后得到修正语义 workflow.实验结果表明,与基于 workflow stream 结构特征的修正算法<sup>[2]</sup>相比,本文的基于 workflow stream 行为特征的语义 workflow 修正算法可以在不存在结构足够相似的匹配 workflow stream 的情况下,使用行为足够相似的匹配 workflow stream 完成检索 workflow 修正;并且得到了整体质量更好的修正语义 workflow 集.本文的 workflow 修正算法为业务过程管理人员为适应新业务需求而进行的工作流变更提供了具有较好参考价值的修正语义 workflow,对提高实际业务 workflow 管理中的 workflow 重用的效率和质量有较大帮助。

在未来的工作中,将致力于将本文的 workflow 修正算法应用于其他领域的业务流程修正;研究更高效的工作流 stream 的行为特征刻画方法和优化的修正操作,进一步提高基于行为特征的语义 workflow 修正算法的效率。

#### References:

- [1] Bergmann R, Gil Y. Similarity assessment and efficient retrieval of semantic workflows. *Information Systems*, 2014,40(1): 115–127.
- [2] Müller G, Bergmann R. Workflow streams: A means for compositional adaptation in process-oriented CBR. In: *Proc. of the Case-Based Reasoning Research and Development*. Springer Int'l Publishing, 2014. 315–329.
- [3] Mantaras RLD, Mcsherry D, Bridge D, *et al.* Retrieval, reuse, revision and retention in case-based reasoning. *Knowledge Engineering Review*, 2005,20(3):215–240.
- [4] Minor M, Bergmann R, Görg S, *et al.* Towards case-based adaptation of workflows. In: *Proc. of the Case-Based Reasoning Research and Development*. Berlin, Heidelberg: Springer-Verlag, 2010. 421–435.
- [5] Minor M, Bergmann R, Görg S, *et al.* Adaptation of cooking instructions following the workflow paradigm. In: *Proc. of the ICCBR 2010 Workshop Proc.* 2010. 199–208.
- [6] Müller G, Bergmann R. Generalization of workflows in process-oriented case-based reasoning. In: *Proc. of the Int'l Flairs Conf.* 2015. 391–396.
- [7] Müller G, Bergmann R. Learning and applying adaptation operators in process-oriented case-based reasoning. In: *Proc. of the Case-Based Reasoning Research and Development*. Springer Int'l Publishing, 2015. 259–274.
- [8] Jin T, Wang JM, Wen LJ. Efficient retrieval of similar workflow models based on behavior. In: *Proc. of the Web Technologies and Applications*. Berlin: Springer-Verlag, 2012. 677–684.
- [9] Zha HP, Wang JM, Wen LJ, *et al.* A workflow net similarity measure based on transition adjacency relations. *Computers in Industry*, 2010,61(5):463–471.
- [10] Bae J, Liu L, Caverlee J, *et al.* Development of distance measures for process mining, discovery and integration. *Int'l Journal of Web Services Research (IJWSR)*, 2007,4(4):1–17.
- [11] Wang JM, HE TF, Wen LJ, *et al.* A behavioral similarity measure between labeled Petri nets based on principal transition sequences. In: *Proc. of the Move to Meaningful Internet Systems (OTM 2010)*. Springer-Verlag, 2010. 394–401.

- [12] Dijkman R, Dumas M, Van Dongen B, *et al.* Similarity of business process models: Metrics and evaluation. *Information Systems*, 2011,36(2):498–516.
- [13] Kunze M, Weidlich M, Weske M. Behavioral similarity—A proper metric. In: Rinderle-Ma S, Toumani F, Wolf K, eds. *Proc of the 9th Int'l Conf. on Business Process Management (BPM 2010)*. Berlin, Heidelberg: Springer-Verlag, 2011. 166–181.
- [14] Sun JY, Gu TL, Wen LJ, *et al.* Similarity algorithm for semantic workflows used in process-oriented case-based reasoning. *Computer Integrated Manufacturing Systems*, 2016,22(2):381–394 (in Chinese with English abstract).
- [15] Song JF, Wen LJ, Wang JM. A similarity measure for process models based on task occurrence relations. *Journal of Computer Research and Development*, 2017,54(4):832–843 (in Chinese with English abstract).
- [16] Dufour-Lussier V, Leber F, Lieber J, *et al.* Automatic case acquisition from texts for process-oriented case-based reasoning. *Information Systems*, 2014,40(1):153–167.
- [17] Kiepuszewski B, Hofstede AHM, Bussler CJ. On structured workflow modelling. In: *Proc. of the Seminal Contributions to Information Systems Engineering*. Berlin, Heidelberg: Springer-Verlag, 1999. 241–256.
- [18] Ullmann JR. An algorithm for subgraph isomorphism. *Journal of the ACM*, 1976,23(1):31–42.
- [19] Bergmann R, Stromer A. MAC/FAC retrieval of semantic workflows. In: *Proc. of the 26th Int'l Florida Artificial Intelligence Research Society Conf.* Menlo Park: AAAI, 2013. 357–362.
- [20] Jin T, Wang JM, Wu NH, *et al.* Efficient and accurate retrieval of business process models through indexing. *LNCS*, 2010,6426:402–409.
- [21] Dongen B, Dijkman R, Mendling J. Measuring similarity between business process models. In: *Proc. of the Int'l Conf. on Advanced Information Systems Engineering*. Springer-Verlag, 2008. 450–464.
- [22] Munkres J. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial & Applied Mathematics*, 1957,5(1):32–38.

## 附中文参考文献:

- [14] 孙晋永,古天龙,闻立杰,钱俊彦.用于面向过程的基于实例推理的语义 workflow 相似性算法. *计算机集成制造系统*,2016,22(2):381–394.
- [15] 宋金凤,闻立杰,王建民.基于任务发生关系的流程模型相似性度量. *计算机研究与发展*,2017,54(4):832–843.



孙晋永(1978—),男,山东枣庄人,博士,讲师,CCF 专业会员,主要研究领域为业务过程管理,知识表示与推理.



钱俊彦(1973—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为软件工程,模型检测,程序验证.



古天龙(1964—),男,博士,教授,博士生导师,主要研究领域为软件工程与形式化方法,知识工程,符号推理.



刘华东(1978—),男,博士生,讲师,主要研究领域为知识工程,符号推理.



闻立杰(1977—),男,博士,副教授,博士生导师,主要研究领域为业务数据管理,大过程数据,业务过程管理, workflow 技术.