IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 12, Issue 5 (Jul. - Aug. 2013), PP 65-76
www.iosrjournals.org

Formalizing Collaborative Software Development Issues: A Collaborative Work Allocation Patterns Analysis Approach

Ch. Srikanth[1], H.Venkateswara Reddy[2]
1 (M.Tech, Cse, Vardhaman college of engineering)
2 (Associate Professor, CSE, Vardhaman College of Engineering)

Abstract: *Currently, workflow machinery is popular to ease the functioning procedure in business in sequence systems (EIS), and it has got the capacity to lessen layout quality and lower product time, improve product price.. This document provides a data mining strategy to deal with the resource allocation issue (RAP) and enhance the efficiency of workflow source executive. Specifically, an Apriori-like algorithm is engaged to discover the regular patterns from the occurrence sign, and organization guidelines are created in accordance with predefined reserve allocation constraints. Consequently, a relationship calculates named lift is utilized to interpret the negatively linked resource allocation guidelines for resource booking. Lastly, the principles are rated as source allocation principles utilizing the trust actions.*
Keywords**:** *workflow, resource allocation, data mining, process mining, association rules.*

I. Introduction

Workflow is now embedded machinery in many enterprises in sequence systems (EIS, e.g. PLM, ERP, CRM, SCM and B2B appliances etc.). Workflow resource allocation serves as an essential link among workflow activities and resources, and it unswervingly determines the implementation quality of the workflow activities [1 3]. Depending on our analysis, the majority of-the resource portion jobs in current workflow organization systems are usually done using a function-based strategy [2, 4, 5]. That is, to break up the workflow possessions (actors) into distinct candidate groups centered on the character and the business qualities. Such resource percentage is relatively coarse grained and might fail in certain scenarios. For instance, in the mechanized businesses, a production progression sheet work may be predetermined to be performed by the assets with the function "process preparation designer". Really, a number of the procedures planning works must be additional assigned to an inferior group of just one or additional capable developers as a substitute of all of the procedure planning developers. Consequently, the current resource allocation procedures might make unsuitable staff projects and the ultimate quality of-the merchandise may endure from this. Accordingly, in some sectors like the production businesses, most of runtime workflow reserve allocation works continue to be done manually by-the managers. Whereas the actions are of great prosperity in some instances, the amount of managers is generally little. That produces a time consuming function to it to spend the workflow possessions manually.

Fortunately sufficient, modern workflow applications generally report the company activities in occurrence logs. These logs commonly include info regarding events talking about an action, an incident, and an instigator [7, 10]. The situation (also called process occurrence) is really a function that has been handled, e.g. a procedure preparation page intend, a compressor style, an NC encoding, etc. As the component of the situation, an action is an example of-a workflow mission. An inventor is the activity is executed by a store (usually RB a person NN In this document, a Procedure describes a workflow pattern of the situation, a Project represents a number of comparable actions, and a Source describes a job performer.

Then, the rules are rated in-a descending collection by their own assurance, and the previous rules are afterwards advised to workflow superintendent at workflow runtime.

The significant benefits of the paper are the following: First, it styles a closed loop workflow platform for-a more sensible and finer grained resource managing. Second, it suggests an alliance rule mining strategy to get the logics among workflow assets as well as the actions, which will aid decision making in resource part.

The balance of the paper is offered as pursues: In Area 2, we style a closed loop workflow structure for optimizing source allocation. Afterwards, we analyze the workflow event designs and their connections in Area 3, and then suggest our mining strategy in Section 4. Finally, we disagree the purpose in workflow resource part in Area 7, and reason this document in Area 8.

II. SE Patterns Opted:

The goal of-the Worldwide Software Advancement for mission Management outline Terminology would be to improve operation of project management function through enhanced international software project managing methods. The GSD outline Language comprises 18 procedure patterns that have already been discovered to be significant within the region of project administration in GSD. The present model of GSD outline Language comprises procedure outline supporting both agile project management and conventional fountain.

GSD outlines are offered in Desk one. The first line includes the title of the design, the 2nd describes the difficulty the design is likely to resolve, and the ultimate column provides outline to the solution of the design. An illustration of the more comprehensive design is really in Table 2.

Table 1. outlines for project management

ID-Name	Problem(s)	Solution outline
01-GSD Strategy	A necessitate of a company level GSD approach.	ist the factors and determination to start GSD based improvement in a business. Produce a long and short term strategy about GSD. Discover the proficiency of distinct websites and create a danger and SWOT evaluation for GSD strategy. Additionally measure the actual prices of GSD.
02-Fuzzy Front End	Unclear how to Assemble merchandise needs internationally from internal and outside clients and the way to	The requirements of various customers may be collected to an international database. In addition it's crucial that you possess the possibility to utilize a discussion forum within the device too as possess the possibility for international access regardless of place and time. Merchandise managers will go through collected needs and make conclusions about them with e.g. designers. If it's recognized in a selection achieving a brand new attribute or demand may be created. Product managers will create a Business model along with a Street Map for-a product including several attributes. These
03-Communicate Early	What is the goal of a GSD project and who are the members of a project?	Arrange kick-off meeting for all relevant members. Present common goal and motivation of this project and present release plan made by Divide and Conquer with Iterations. Also present responsibilities made by Work Allocation. Present used Common Processes and Common Repositories and Tools. Organize leisure activities for teams to improve team spirit.
04-Divide and Conquer with Iterations		See an example below (Table 2).
05-Key Roles in Sites	Difficult to know who to contact in different sites with your questions.	A manager will have discussions with website managers or other administrators about group members before final choices. Also needed functions may be created in every website (e.g. Website task Manager, Architect, IT Assistance, Quality confidence etc.) The primary website individual is really in a top place as well as the individuals from some other websites will look after the duties, jobs and issues within their websites. Release the entire endeavor business with functions for each website to enhance communication. One individual might have many functions in a
06-Communication Tools	Need of communication tools can also vary among sites.	Have typical and reliable communication approaches and resources in every website. Use various tools at-the similar period as internet assembly to demonstrate tips and job data, conference phones to possess great seem and talk application to discuss in composed form if there are issues to comprehend e.g. English-language used in additional websites. Also train and move job
07-Common Repositories and Tools	Separate Excel files are complicated to manage and project data is complicated to find, manage and synchronize among many sites.	Give a frequent Application Life-cycle Management (ALM) tools for-all job artefacts (files, source code, bugs, guidelines etc.). ALM provides nearly realtime traceability, coverage, creation and entry to required advice etc. for all consumers in various sites. It may be applied as just one instrument or it may become a number of various resources which is incorporated with every other. ALM resources may comprise means to help procedure based on the organisation's procedures and development approaches (state models, process templates, workflows). Use various
08-Work Allocation	Work needs to be shared among sites with various criteria.	Learn what the GSD Strategy is in-your business and check advice of individuals in every single site with aid of site supervisors. Make a choice about department of function between sites in accordance with a business's GSD Strategy and the aforementioned evaluation.
09-Architect-ural Work Allocation	Work needs to be shared among sites with architectural criteria.	Check architectural evaluation of the strategy and goods which website will be in charge of keeping and raising understanding in certain architectural area. Architectural area may still be a complete subsystem or part of-a subsystem.
10-Phase-Based Work Allocation	Work needs to be mutual between sites with phased-based criteria.	Examine how stage- based function allocation may be produced. Also check which website is perhaps in charge of keeping and raising understanding in some period-based area e.g. testing or requirements engineering in a particular merchandise area.

11-Feature-Based Work Allocation	Work needs to be shared among sites with feature- based criteria.	Assess the GSD Strategy how feature- based work allocation strategy is described. Form a number of people from various sites if desired, to understand the attributes.
12-Use Common Processes	Dissimilar processes and templates at different sites make communication	Choose frequent upper level processes and allow local processes if they do not origin problems with upper level processes.
13-Iteration Planning	Persons do not know what kinds of features are desired for a GSD project and what the existing goal is.	Manager will present prioritized attributes and additional jobs. Project members may participate in-a planning meeting either individually o-r by Communication Resources. The task members will estimate quantity of function for jobs and attributes. If desired, more comprehensive discussion can be organized in websites with participants' mother language. Ultimately of preparation, meeting the set of jobs and chosen characteristics is
14-Multi Level Daily Meetings	Problems to have a daily frequent meeting with all members with dissimilar time zones.	Organize many daily conferences and organize another daily o-r weekly conference between project supervisors from various websites to trade info regarding the effects of daily conferences. With foreigners, written logs may be one option to make certain that communication communications are recognized correctly in every website. Pick the same working period for
15-Iteration Review	It's complex to know what the status of a mission is and the feedback loop is	Assess the job status by a demonstration and current leads to all stakeholders and applicable people from various websites. Collect trade requests and opinions for additional measures for both procedure and merchandise. Make regular deliveries to enhance awareness of the standing
16-Organize Knowledge Transfer	It's complex to transfer a huge amount of knowledge to new or practiced developers of different sites.	Make certain that there's a product knowledge repository accessible for task members. Teach the merchandise and get members additionally to use. Specs with use-cases will be shown within the Technology Planning conference or individual conferences. Also earlier customer paperwork and demonstration will be shown in certain instances. Crucial Functions in Websites network is going to be used by attempting to get options for difficulties. Use regular or longer appointments to enrich knowledge exchange and be certain we have great communication stations between
17-Manage Competence	It's difficult to know what the competency of each project member is.	Produce a competence database for collecting info of members' competence levels at various sites. Define standards and competence levels for them. Determine the regions of competency you need to track.
18-Notice Cultural Differences	Certain methods are suitable in one nation's culture and might not be suitable in another.	Increase the understanding of your team countries' culture for team people. Use ambassadors, site visits and liaisons, if possible. When you're employing GSD Strategy and Function Allocation discover ethnic differences. Use Common Procedures. Use Common Databases and Conversation Resources and Resources. Enable local strategies in procedures, tools, meeting procedures etc. to reduce problems with ethnic

Table 2. **An illustration of GSD pattern.**

Name:	GSD 04 Divide and overcome with Iterations
Problem:	One big project arrangement is a risk in dispersed development and long feedback loops.
Solution:	Implement the subsequent actions: • Plan many iterations to illustrate the project plan • Develop new submission architecture and module structure during first iterations, if needed • Explore the biggest risks (e.g. new technologies) in the establishment of a project • The length of iteration can be e.g. 2-4 weeks to progress control and visibility.
Resulting Context:	• Iterations improve the visibility of a project and enthusiasm of project members • Iterations make it easier to control a project when you gash the whole project into many controllable parts • Administration work is improved with many iterations

III. A closed-loop workflow outline for resource portion

Our work is pedestal on a National resistance Project named Agile Process investigate System (APPS) for a large radar-manufacturing conglomerate [13] in Nanjing, Jiangsu, China. APPS is a process-aware in sequence system, and it applies a workflow component to handle the works of CAX units (e.g. CAD, CAM, CAPP, etc.). This workflow component manages the possessions (performers/actors) using a closed-loop approach. The framework of the appear is illustrated in Fig. 1.

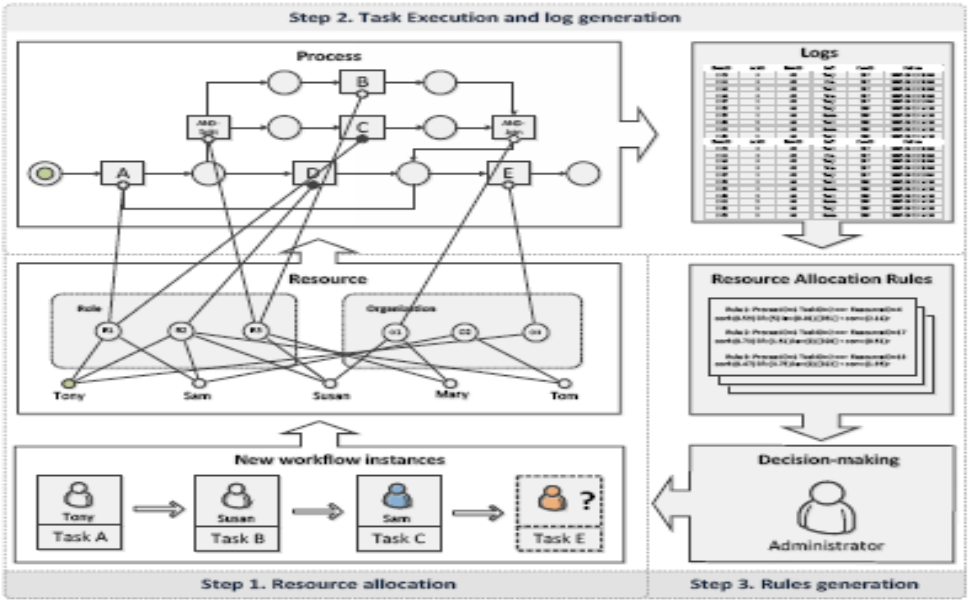

Fig. 1 **outline of the approach.**

The closed loop workflow resource allocation strategy chiefly comprises three steps:

Action 1: The delivery history of-the workflow actions is documented in-a deal record known as workflow record

Action 2: The program uses an Apriori-like organization statute mining algorithm to pull the source portion knowledge from the workflow confirmation.

Action 3: the system automatically suggests the manager using a default source and other suitable candidates based on the mined association rules, Once a brand new workflow action is began. The workflow superintendent might only endorse the default task or pick another resource within the applicant register for the function considering the world.

From workflow monitor to Resource portion Rules

Our aim is to distill resource portion guidelines with large prediction precision away from the work-flow event record. A workflow occurrence usually comprises three main types of details: the workflow job info, the workflow procedure details as well as the resource details. The organization rule including these three proportions without continued predicates cascade in the multidimensional organization rules mining domain[14].

Models and units in workflow

Workflow replica

To illustrate, we use a creation planning process p_{66}, and the workflow illustration regular to the process is represented in Fig. 2.

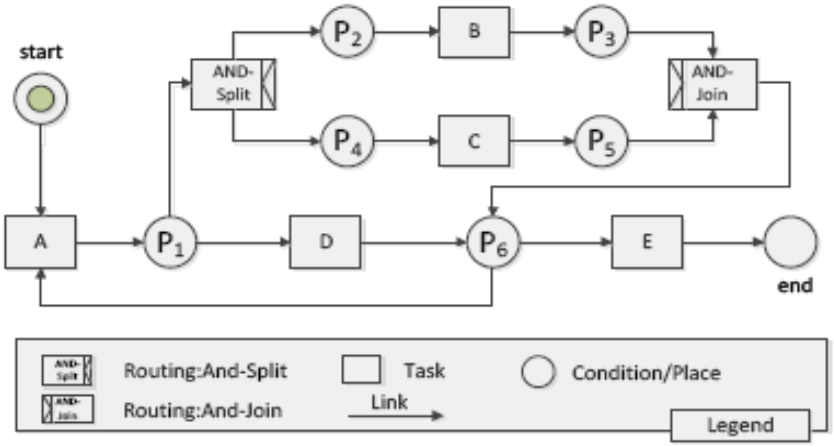

Fig. 2 A illustration process P_{66} modeled using WF-NET

Fig. 2 shows a simplified workflow sequence p_{66} modeled with WF-NET[4]. This process comprises five tasks (A, B, C, D, and E), a similar routing (the AND-Split and the AND-Join), and two perceptive routings (OR-Split P_1 and OR-Join P_6).

Table 1 A element of the Workflow Log

Table 1 is an event log illustration dependable with the process P_{66} . This sample mainly includes some entities of WfMS: resource and task, case, and process, the "EventID" is the individuality of the log and the "CaseID" referred to the personality of the instances of P_{66} .

EventID	ActID	FlowID	Staff	CaseID	SetDate
5313	1	66	Tony	203	2007-10-15 21:06
5314	1	66	Sam	204	2007-10-15 21:09
5315	4	66	Mary	203	2007-10-15 21:10
5316	1	66	Sam	205	2007-10-15 21:11
5317	1	66	Tony	205	2007-10-15 21:13
5318	1	66	Tony	206	2007-10-16 13:46
5319	2	66	Tom	204	2007-10-16 13:47
5320	1	66	Sam	203	2007-10-16 13:49
5321	4	66	Susan	203	2007-10-16 13:50
5322	1	66	Mary	206	2007-10-16 13:51
5323	2	66	Mary	204	2007-10-16 13:52
5324	3	66	Tony	204	2007-10-16 13:53
5325	3	66	Tom	204	2007-10-16 13:54
5326	1	66	Sam	206	2007-10-16 13:55
5327	2	66	Tom	205	2007-10-16 13:56
5328	5	66	Susan	203	2007-10-16 14:02
5329	4	66	Sam	206	2007-10-23 14:04
5330	2	66	Mary	205	2007-10-23 14:05
5331	1	66	Susan	204	2007-10-23 14:06
5332	1	66	Mary	206	2007-10-23 14:08
5333	3	66	Sam	205	2007-10-23 14:10
5334	3	66	Susan	205	2007-10-23 14:13
5335	2	66	Tony	206	2007-10-23 14:14
5336	3	66	Susan	206	2007-10-23 15:31
5337	3	66	Tom	206	2007-10-23 15:33
5338	4	66	Sam	204	2007-10-23 15:37
5339	3	66	Susan	206	2007-10-23 15:38
5340	5	66	Tom	205	2007-10-23 15:42
5341	5	66	Susan	205	2007-10-23 15:43
5342	5	66	Tom	204	2007-10-23 15:44
5343	5	66	Susan	206	2007-10-23 15:47
5344	5	66	Tom	205	2007-10-23 15:50

This procedure is made in a Petrinet-like model known as WF NET. Organizations in this plans are : routing, and procedure, endeavor, etc [4, 15]. First, in Job A, the device instantly searches the database for comparable cases. The procedure will be suggested to Job D, and the custom will change them and obtain the case files, supposing that there are cases meeting certain requirements. The job could be approved through a similar routing to Task D and both Task W, if there's no well suited case, and fresh design jobs may be designated to corresponding developers for Task W and co - designer for Task D. The file may probably be aged in Task E, when both the Job W and D are concluded and the entire layout process finishes. Reminder that, we don't regard as iteration routings, like the stripe from area to Job A. A workflow log creates as the works perform from one measure to a different consistent with-the manage circulation of-the procedure in Fig. 2.

Neither do we regard as the order of the activities corresponding to various instances. For instance, as

is demonstrated in Desk table 1 suppose that a workflow log consists the following workflow operation events: $e_1:(p_1,t_2,r_4)$, $e_2:(p_2,t_1,r_3)$, $e_3:(p_1,t_2,r_9)$, $e_4:(p_2,t_1,r_7)$, $e_5:(p_3,t_3,r_{10})$. This progression can be divided into movement set view according to progression id and task id: Activity Set 1 $(P_1,t_2):(e_1,e_3)$; motion Set 2 $(P_2,t_1):(e_2,e_4)$; motion Set 3 $(P_3,t_3):(e_5)$, where each Activity matched to a same pair of process and task. Thus, we get a representation view of the sample of the workflow record in Table 2:

Table 2A representation view of the workflow log

	CaseID	Log events
1		(A, Tony), (B, Susan), (C, Tom), (E, Mary)
2		(A, Tony), (D, Jim), (E, Mary)
3		(A, Tony), (B, Susan), (C, Tom), (E, Mary)
4		(A, Tony), (D, Jim), (E, Mary)
5		(A, Tony), (B, Sam), (C, Sam), (E, Tony)
6		(A, Jim), (B, Susan), (C, Sam), (E, Tony)

For convenience, we define some models used in the mining process by adopting some notions defined in Ref. [15].

Definition1. (Workflow procedure)

A procedure indicates the working errands and the orders in which this must be done. Let $p=\{p_1,p_2,.....p_{NP}\}$ be a set of procedures, where P_i is a process. A task is an tiny unit of a process, let T be a set of tasks, $i\in\{1,2,....n_p\}$. Let R be the position of performers/originators (i.e., staffs, resources, or means), $R=\{r_1,r_2,...r_{nr}\}$

IV. Workflow log

To deal with belongings is the main aim of a work-flow system, where in fact the jobs of comparable cases are structured in the exact same ways, specifically workflow procedures. Put simply, a situation is an example of some procedure. As the circumstances run in the work-flow program the trade events create.

Therefore, we conceptual the event record for a collection of quadruples: (case, project, reserve, timestamp) The tuple is indicated by The tuple resource performs the project of an example of procedure in a particular time. Within this document, we give attention to who do what job in which procedure and don't treatment much a propos the performance instance and the series of the jobs (cases). What passions us is consequently the co occurrence of procedures, jobs, and sources. As-a case is an example of-a process, we could certainly get the procedure details with case details. Let us represent the sets $P=\{p_1,p_2,.....p_K\}$ $T=\{t_1,t_2,...t_M\}$ $R=\{r_1,r_2,...r_N\}$ to be the sets of K procedure, M tasks and N possessions in log L. With the log information pretreatment (omit the time in sequence and get the process in sequence from the cases), we can interpret each quadruple to a triple of (*procedure, task, resource)* .

Consider a sample of the workflow operation log in our workflow managing system shown in Table 1. Typically, the log enclosed thousands of records, where each confirmation refers to a certain workflow motion. An activity is an implementation composed of a procedure, a task, and a resource, perhaps a staff.

Definition 2. (Event log)

Let E = P x T x R be the set of (potential) events, an event $e_s:(P_i,t_j,r_k)$ logs a workflow motion comprised of a procedure P_i, a task t_j and a resource r_k (the inventor). $C=E^*$ is the set of potential event sequences (traces recounting a case). $L\in B(C)$ is an event log, here B(C) is the position of all bags (multi-sets) over C. Each constituent of L denotes a case.

It means that the task *t* of procedure *p* is executed by resource *r*. In WfMS, $R=\{r_1,r_2,...r_{nr}\}$ and n_r is the number of the reserve units, n_r = |R|.

3.1. Resource portion rules representation

In this work, there is a multidimensional information warehouse with four consistent relations as is shown below:

- Workflow_log (EventID, ProcessID, TaskID, ResourceID, EventType, CaseID),
- Process (ProcessID, ProcessName,),
- Task (TaskID, TaskName, ProcessID, TaskType, Desription,),
- Resource (ResourceID, ResourceName, HasOrgEntity, HasRoleEntity),

Where Process, Task, Resource are three measurement tables. These tables are associated to the Workflow_log table via three keys: ProcessID, TaskID and ResourceID. The
Correlated star representation of our store is depicted in Fig. 3.

The requirements must be met by the project for mining brought out in [15] before method mining. The record information is preprocessed to get certified for process mining, these preprocessing include revising or removing the imperfect or unsound data. Defective or unsound data chiefly submit to those incomplete data, which are missing the crucial data items like the action, method, situation or originator. Accordingly, before mining, some formulations are crucial.

The Source is the title of the originators. The record section comprises 33 occasions, requires 5 actions, 4 cases, 5 assets and a procedure.

Rather than hunting on just one feature like procedure, we must operate through multidimensional characteristics including job, procedure and resource, as an itemset treating each pair. We utilize the multidimensional data demonstrated in-the star schema in Fig. three to build a data cube [17 19]. The generalization of team by, rollup and cross-tablature thoughts would be to comprehensive the measurements. In Fig. 3, it's a 3 dimensional data cube, along with the conventional GROUP BY creates the 3 dimensional data cube primary.

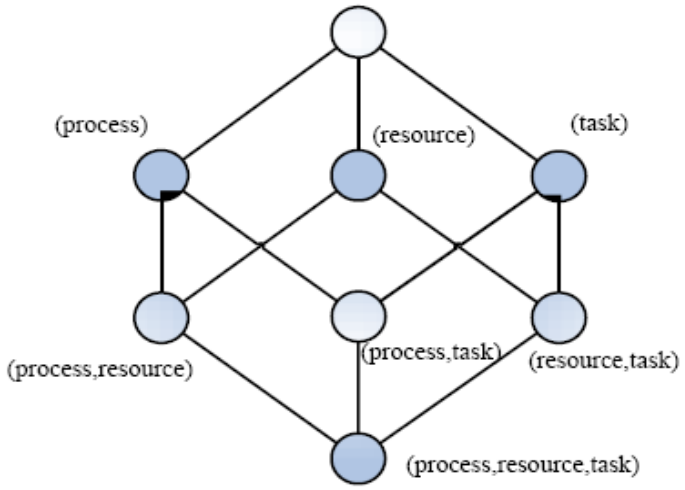

Fig. 3 The 3-D data cube[14].

Information cubes are well suited for removal multidimensional organization rules. Fig. 3 shows the lattice of cuboids essential a data cube for the dimensions procedure, task and resource. An alliance rule has the form like *LHS* ^ *RHS*, that is, from Left Hand Side (LHS) to Right Hand Side (RHS). By using the information cube, we may get some different multidimensional rules.

Let us now see an instance of a single frequent 3-itemset $F_3 : \{P_1, t_1, r_9\}$,which is resultant from the event log via the algorithm in section 4.1. The nonempty F_3 are $\{P_1, t_1\}$, $\{t_1, r_9\}$, $\{P_1, r_9\}$, {p}, {t}, $\{r_9\}$. Thus, we can get the involvement rule in different forms:

Whereas several of the rules are of no help to reserve allocations, e.g., the rules in the form of $t_1 \Rightarrow P_1$ means that the task t_1 of procedure P_1 is frequent performed in the scheme. Hence, we have to use the dimension/level limitation [20] to filter out the rules with little notice.

Definition 3. (reserve Allocation Rule Constraint)

For an motion of Task Y in Process X, and the Resource Z, our explore objective in this paper is to find the reserve distribution rule as follows called PTR (procedure, task to resource) metarule:

If we filter the rules using limitation in Definition 3, then only the rule p a t ^ r_9 is capable output. We can get a register of the rules by iterating this step to all of the normal itemsets in the event log. "Find the execution of what task may support the working frequency of the possessions in the same case (the occurrence of a process)" is an organization rules mining query, which can be articulated in a data mining query language (DMQL) as follows:

This mining uncertainty allow the generation of connection rules in the form as below:

The rules indicate that if a work motion of the Task 1 in Process 9 is to be performed, there is a 59% probability that the work will be execute by Resource 3, Tom. A further suggestion of the rule is that, 3.2% of all the workflow events fulfilled all the criteria, and the lift calculate of this rule is 5, indicating it a absolutely correlated organization rule.

Generate, annotate and rank: A three-stage move toward to mine resource allocation rules

We have initiated some basic terms in reserve allocation rules mining in preceding subsections. In order to quarry the multidimensional connection rules from the event logs, in this paragraph, we present a three-stage loom to get the useful rules.

Stage 1. Generated raw reserve allocation rules: Find all regular 3-itemsets, and generate resource distribution rules using the rule limitation in Definition 3, and by definition, each of the itemsets must satisfy the minimum support and least confidence.

Stage 2. Annotate the rules by Negative connection Annotation algorithm.

Stage 3. Make a rule progression by assurance of the rules using resource portion rules sorting method.

4.1. Frequent resource portion rules generation

In connection rules mining domain, an itemset *I* is normal only if its *support* value satisfies the minimum support threshold *minsup.* The term *support* here is also referred to as *relative support,* and it indicates the occurrence frequency of the itemsets[14]. For association rules mining in the form of $P\Lambda t \Rightarrow r$ we define the term support as:

$$\sup(p\Lambda t \Rightarrow r) = \frac{count(p,t,r)}{count(L)}$$

As is shown in Eq.(2), the support measure is the proportion of transactions in *L* that contain the itemset (p, t, *r*), The purpose *count(L)* returns the number of records in the log, and the *count*(p, t, s) returns the count of event logs equivalent to the process *p* , task *t* and resource *r* .

Frequent itemset removal leads to the discovery of associations and correlations between items in large transactional information sets. However, this can be a time-consuming method. In this paper, we use the Apriori algorithm to find common patterns. Apriori is a classical algorithm projected by R. Agrawal and R. Srikant in 1994 for mining organization rules, and is proved to be competent and scalable for both reproduction and real world data sets[11, 12]. The high presentation of this algorithm is based on the priori information that all nonempty subsets of a common itemset must also be common, and here we use its contraposition.

We concern the Apriori algorithm along with the "Resource Allocation Constraint". According to Definition the regular itemset should be 3-dimentional, and the regular itemsets must satisfy *min_sup* threshold. We can get the mining algorithm below:

Mining Multidimensional organization rules [14] from workflow event logs.

Algorithm: Frequent-pattern generation. Find common itemsets using an iterative level-wise loom based on Apriori candidate production.

Input:

- *L,* the workflow event log;
- *Min_sup,* the minimum sustain count threshold. **Output:** *F* , frequent 3-itemsets in *L.*

Method:

(1) F = *find _ frequent* _1 — *itemsets(L);*

(2) **for** $(k = 2; F_{k-1} \neq \phi; k++)\{$

(3) *Ck = apriori _gen* F_{k-1})

(4) **for each** transaction *t e L* {//scan L for counts

(5) C_t = *subset*(C_k *t*) //get the subsets of t that are candidates

(6) **for each** candidate *c e* C_t

(7) *c.cuu* ++ ;

(8) }

(9) $F_k = \{c \in C_k \,|\, ccount \supseteq \min\_ s\}$

(10) }

(11) return $L = \bigcup_3 L_3$; //Generate frequent 3-itemsets with dimension constraints.

Procedure *apriori _ gen(* F_{k-1} ; *frequent(k* -1) - *itemsets)*

(1) **for each** itemset $l_1 \in F_{k-1}$

(2) **for each** itemset $l_2 \in F_{k-1}$

(3)if

(5))

(6) **then** {

(7) $c = l_1 \otimes l_2$;//joint step: generate candidates

(8) **if** has_infrequent_subset(c , F_{k-1}) then

(9) **delete** *c* ; //prune step: remove unfruitful candidate

(10) **else add** *c* to ;

(11) }

(12) *return* **;**

Procedure *has in frequentsubset* (c : *contender k — itemset*; F_{k-1} : *frequent(k* -1) - *itemsets*); // use the prior information

(1) **for each** (k -1) -subset s of *c*

(2) **if** $s \in F_{k-1}$ **then**

(3) **return** TRUE;

(4) **return FALSE;**

Once that the common 3-itemsets from the log have been originated, it is straightforward to produce strong rules from them. *Strong rules* are those who both satisfy *least support threshold (min_sup)* and *minimum assurance threshold (min_conf)* . The rule $p\Lambda t \Rightarrow r$ has *confidence c* in the communication log set L, where *c* is the proportion of transactions in L containing $p\Lambda t$ that also contain *r*. It is a conditional prospect

$$\text{Confidence } (p\Lambda t \Rightarrow r) = P(r \mid p\Lambda t) \frac{Support\_count(p\Lambda t\Lambda r)}{Support\_count(p\Lambda t)}$$

To convert the common itemsets into strong resource allocation rules, we use the check in Definition 3to confine the measurement and form of the mined rules, and we may get a list of these "capable in form" rules below:

Rule 1: process=8 task=1 655 ==> resource=19 655 conf:(1)
Rule 2: process=7 task=1 206 ==> resource=17 199 conf:(0.97)
Rule 3: process=5 task=8 296 ==> resource=4 276 conf:(0.93)

In the preceding sections, we argue the method of finding the frequent executors for the workflow tasks. However, the rules mined with the support-confidence structure discussed beyond may disclose some not so appealing event relationships[21, 22]. Let us study the attached rules:

Rule 1: ProcessID=1 TaskID=2 ==> ResourceID=4 conf:(0.59) [support_count=967]
Rule 2: ProcessID=1 TaskID=2 ==> ResourceID=17 conf:(0.20) [support_count=328]
Rule 3: ProcessID=1 TaskID=2 ==> ResourceID=13 conf:(0.13) [support_count=213]

As illustrated in the list, all of the rules are above the support/confidence threshold. However, Rule 2 would be misleading when $P(p_1\Lambda t_2 \Rightarrow r_{17}) = 0.20 \le p(r_{17}) = 0.40$. Therefore, by characterization, $LHS(p_1\Lambda t_2)$ and $RHS(r_{17})$ are really negatively correlated as the existence of LHS essentially decreases.

Lift is a connection measure used to find out unexciting rules. A rule *LHS* $\Rightarrow$ *RHS* is negatively correlated if *lift (LHS* $\Rightarrow$ *RHS*) < 1, else, it is positively connected. In this paper, we interpret the negatively correlated rules and suggest them to the administrators as substitute resource candidates, along with the positive ones.

The negatively association annotation indicates that, although it is suitable to assign the annotated resources to the workflow action, it is better to keep them in reserve for their major works.

The negatively-correlated-rule-annotation algorithm is as follows:

Annotate pessimistically correlated resource distribution rules in the strong rules

Algorithm: Negative correlated connection rules annotation.

explain the rules with negative correlation.

Input:

■ S, the candidate strong reserve allocation rule set; **Output:** *AR* , resource portion rule with negative connection annotation;

Method:

(1) **for each** resource allocation rule $rl \in S$ { //Scan *L* for counts

(2) *if* **lift(** rl **)** *< 1 then*

(3) **annotate** rl as negative correlated; // Annotate the negative correlated rules

(4) **else add** rl to *AR* ;

Rules confidence position: sort rules by confidence

In association regulation mining area, a foremost method to sort a compilation of association rules is the most-confident assortment method[24, 25]. The most-confident rule assortment method always chooses the highest assurance among all the association rules whose sustain value is above the *min sup* threshold. Hence, we use the assurance measure to sort the resulting rules to produce the resource allocation rules list for assessment support.

When the PTR rules are produced, the rules are then divided into dissimilar sets by their LHS. Suppose that for a specific rule set with the LHS (p_3 A t_6), the mined strong PTR rules are:

Rule 1: ProcessID=3 TaskID=6 ==> ResourceID=7 conf:(0.26)
[support_count=426]

Rule 2: ProcessID=3 TaskID=6 ==> ResourceID=11 conf:(0.54) [support_count=885]

Rule 3: ProcessID=3 TaskID=6 ==> ResourceID=13 conf:(0.10) [support_count= 164] In this example, the confidence values of Rule 1, Rule 2, and Rule 3 are 0.26, 0.54 and 0.10, respectively. Then we get the ranked rule list by the confidence measure in descendant order:

Rule 2: ProcessID=3 TaskID=6 ==> ResourceID=11 conf:(0.54) [support_count=88 5]

Rule 1: ProcessID=3 TaskID=6 ==> ResourceID=7 conf:(0.26)
[support_count=426]

Rule 3: ProcessID=3 TaskID=6 ==> ResourceID=13 conf:(0.10) [support_count=164]

With most-confident assortment method, the system then automatically chooses the reserve r_{11} from rule 2 as defaulting recommendation for the administrator. Note that in our loom, the system will also recommend the resources optional by rest of the list, r_7 and r_{13} (from Rule 1 and Rule 3) as alternatives. For N different test cases, let C be the number of accurate predictions, then the resource forecast accuracy of the activity ($p_3 \Lambda t_6$) is:

$$precision = \frac{C}{N}$$

The rationale of most-confident assortment method is that the testing information will share the same characteristics as the training information [25, 26]. Thus, if a rule has a high assurance in the training data, then this rule would also show a high correctness in the testing data.

Experiment and evaluation

Experiment setup

Our effort is based on the workflow history information from a PDM system named KM PDM (http://www.kmsoft.com.cn/Contents-119.aspx) deployed in a huge electronic manufacturing enterprise[13] in Nanjing, China. We significance the event information of 10 processes from the KM PDM database using SQL queries. Given the workflow log information, the first step is to clean the raw data. We clean out noise logs with no originators and those logs achieved automatically or allocated to originators at design-time (The subsistence of these event logs will not help us in removal the run-time resource portion rules). Finally, we get a log with 75934 items.

Training information overview

Table 3 shows the effecting frequency counts of the preparation log, each column of the table shows the task progression number in the proces, and each row communicate to a process. As we can see, the columns of the table are the progression, and the rows signify the # of the tasks in processes, and the numbers in the cells are the occurrence counts of the corresponding process and task. There are 10 processes and 141 tasks in the preparation dataset.

Process/task Task										0
	17	020	81	80	09	42	06	55	55	118
2	78	68	109	20	61	79	89	69	08	48
3	35	84	64	98	62	08	53	77	22	83
4	40	86	71	35	02	35	67	73	76	98
5	78	82	22	83	11	16	18	30	10	42
6	15	97	70	90	24	21	77	90	62	76
7	07	75	201	07	5	84	95	15	63	65
8	46	04	189	41	96	92	28	53	37	19
9	58	41	13	69	73	05	29	04	14	04
10	44	26	81	67		66	78	60	76	61
11	095	07	83	59		21	98	47	87	59
12	58	88	79	15			45	62	98	0
13	168	24	29	19			13		18	0
14	10	77	98				28		07	0
15	85	56							18	0
16	57	378							90	0

Table 3 Process-task distribution in training data

Fig. 4 demonstrates the comparative frequency circulation of the processes of the workflow logs. The X-axes are the # of the resources, actions and processes; the vertical axes symbolized the occurrence occurrence or relative frequency.

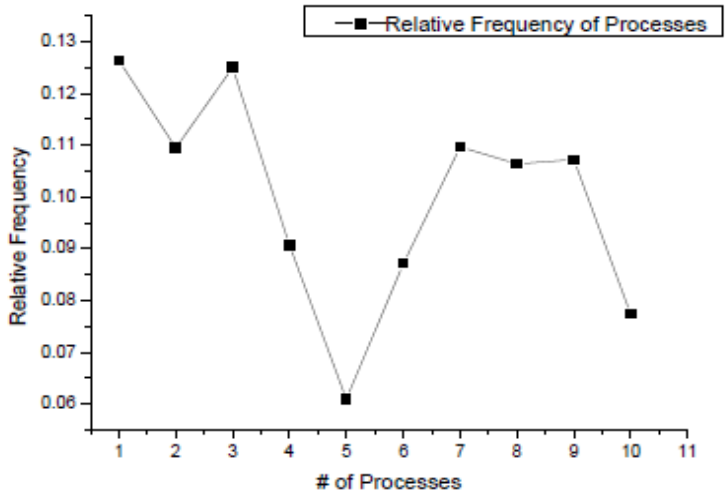

Fig. 4 Process division in the training data

Experiment results

After the provision made above, we use the Apriori algorithm to produce association rules from the workflow log.

After we get the huge itemsets, we process the information with the 3-stage method referred in Section 4. For motion act ($p_6 \Lambda t_9$), we find in the log the rule list as:

Stage 1. Generate the association rules: With these large itemsets we can get the association rules above *min_sup* threshold and under the resource allocation constraint in Definition 3.

Rule 10: progression =6 thread=9 1053 ==> result=20 7conf:(0.0066)
Rule 11: progression =6 thread =9 1053 ==> result =7 6 conf:(0.0057)
Rule 12: progression =6 thread =9 1053 ==> result =10 6conf:(0.0057)
Rule 13: progression =6 thread =9 1053 ==> result =18 6conf:(0.0057)

Stage 2. Annotate the rules: interpret the negatively associated rules with mark.

Stage 3. Sort the rules in precedence: sort the rules with the assurance measure.

The functionality of-the proposed classifier according to the most assured choice method is acceptable compared with these in [2, 28]. Though, the complete forecast precision of approximately 60% also means that regarding 40% of every one of the system assigned workflow activities require manual reassignments. And So, the guidelines with greatest precision for the screening data aren't usually the top alternative. Take actl05 (p a /9) for illustration, Rule one is of a confidence 51%, as well as the amount of best 3 absolutely linked rules reaches around much as 80.63%.

Consequently, instead of recommending one best forecast for every class of workflow actions, the program also urge other strong resource portion guidelines to the workflow manager as applicants: when the assets with high confidence are inaccessible at the second, the remaining applicants (including the annotated assets) in the checklist could be the choices.

Additionally, with the aid of the negatively linked rules observations, the administrators may get a holistic perspective of the sources' work priorities.

V. Conclusions and future work

We've offered a decision making strategy utilizing data mining knowledge to make suggestions to workflow initiators. Feasibility assessment using a explore study indicates the offered strategy could be helpful in sustaining workflow resource portion.

Then we converse the benefits and limits of the technique. Alongside the administrators' comprehension of the workload of the assets, and expert information to distinct product design jobs, our strategy may nicely manage the majority of the source allocation issues in PAISs. Our potential work contains two primary components: (1) examine another machine learning methods like inductive knowledge programming (ILP) with our current method to discover even more effective and powerful methods. (2) see the source distribution rules from various directorial levels and measurements (e.g. the functions and the business units).

References

[1] L. Ly, S. Rinderle, P. Dadam, and M. Reichert, "Mining Staff Assignment Rules from Event-Based Data," in *Business Process Management Workshops*, ed, 2006, pp. 177-190.

[2] Y. Liu, J. Wang, Y. Yang, and J. Sun, "A semi-automatic approach for workflow staff assignment," *Computers in Industry*, vol. 59, pp. 463-476, 2008.

[3] Z. Huang, W. M. P. van der Aalst, X. Lu, and H. Duan, "An adaptive work distribution mechanism based on reinforcement learning," *Expert Systems with Applications*, vol. 37, pp. 7533-7541, 2010.

[4] W. van der Aalst and K. van Hee, *Workflow Management: Models, Methods, and Systems*vol. 1: The MIT Press, 2004.

[5] W. M. P. van der Aalst, A. H. M. ter Hofstede, B. Kiepuszewski, and A. P. Barros, "Workflow Patterns," *Distributed and Parallel Databases*, vol. 14, pp. 5-51, 2003.

[6] N. Russell, W. M. P. van der Aalst, A. H. M. ter Hofstede, and D. Edmond, "Workflow Resource Patterns: Identification, Representation and Tool Support," in *Advanced Information Systems Engineering*. vol. 3520, O. Pastor and J. Falcão e Cunha, Eds., ed: Springer Berlin / Heidelberg, 2005, pp. 216-232.

[7] J. E. Cook and A. L. Wolf, "Discovering models of software processes from event-based data," *ACM Trans. Softw. Eng. Methodol.*, vol. 7, pp. 215-249, 1998.

[8] W. M. P. van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters, "Workflow mining: A survey of issues and approaches," *Data & Knowledge Engineering*, vol. 47, pp. 237-267, 2003.

[9] W. M. P. van der Aalst and A. J. M. M. Weijters, "Process mining: a research agenda," *Computers in Industry*, vol. 53, pp. 231-244, 2004.

[10] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song, and H. M. W. Verbeek, "Business process mining: An industrial application," *Information Systems*, vol. 32, pp. 713-732, 2007.

[11] R. Agrawal and R. Srikan, "Fast Algorithms for Mining Association Rules " in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994, pp. 487-499.

[12] Z. Zheng, R. Kohavi, and L. Mason, "Real world performance of association rule algorithms," presented at the Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, California, 2001.

[13] (2009). *Nanjing Research Institute of Electronics Technology (Home page)*. Available: http://www.nriet.com/

[14] J. Han and M. Kamber, *Data Mining:Concepts and Techniques*: Morgan Kaufmann, 2006.

[15] W. M. P. van der Aalst, Song, M., "Mining Social Networks: Uncovering Interaction Patterns in Business Processes," *Business Process Management: Second International*