# **Process Mining**

Luigi Pontieri Istituto di Calcolo e Reti ad Alte Prestazioni ICAR-CNR Via Bucci 41c, Rende (CS) pontieri@icar.cnr.it



# Argomenti



- Caratteristiche generali delle tecniche di Process Mining (PM)
  - II PM come approccio all'analisi (ex-post) di processi organizzativi
  - Caratteristiche dei processi e dei dati (log) oggetto dell'analisi
  - Obiettivi, potenzialità e problematiche correlate al Process Mining
  - Inquadramento del PM nel ciclo vita dei processi organizzativi
- Classificazione degli approcci di Process Mining
  - Analysis perspectives: Control-flow, Case, Performances
  - Tasks: Discovery, Extension, Conformance testing
- Approfondimento su alcune tecniche e di workflow discovery
  - Induzione di Control Flow graphs: algoritmo di base
  - Uno sguardo ad alcuni approcci classici (α-algorithm, HeuristicMiner, Multi-phase, Fuzzy)



# Argomenti (2)

#### Valutazione e validazione dei modelli (scoperti o pre-esistenti)

- Conformance Checking
- Log-based property verification

#### Altri task e metodi di PM

- Induzione di modelli organizzativi e di social networks (cenni)
- Tecniche clustering-based per la scoperta di schemi di processo gerarchici/tassonomici
- Tecniche per l'estensione di un modello di processo
- Ulteriori linee di sviluppo del PM
  - Scoperta di istanze di esecuzione anomale
  - Integrazione del PM con ontologie di processo e di dominio

# Organizzazione



#### Lezioni

- Teoria di base
- Strumenti SW per il Process Mining
  - Esempi di uso della suite open-source ProM
- Casi di studio

#### Esercitazioni

- Esercizi sui concetti appresi nelle lezioni
- Analisi di alcuni dataset di esempio con ProM

## Materiale didattico



- Lezioni (slide MS PowerPoint):
  - http://www.icar.cnr.it/pontieri/didattica/PM/slides/
- Riferimenti bibliografici
  - I. Witten, E. Frank, Data Mining: Practical Machine Learning Tools with Java Implementation. Morgan Kaufman, 1999
  - Una serie di articoli scientifici disponibili all'indirizzo <u>http://www.icar.cnr.it/pontieri/didattica/PM/papers/</u>



## Outline

#### Part I – Introduction to Process Mining

- Context, motivation and goal
- General characteristics of the analyzed processes and logs
- Classification of Process Mining approaches
- Part II Workflow discovery
  - Induction of basic Control Flow graphs
  - $\Box$  Other techniques ( $\alpha$ -algorithm, Heuristic Miner, Fuzzy mining)
- Part III Beyond control-flow mining
  - Organizational mining
  - Social net discovery
  - Extension of workflow models
- Part IV Evaluation and validation of discovered models
  - Conformance Check
  - Log-based property verification
- Part V Clustering-based Process Mining
  - Discovery of hierarchical workflow models
  - Discovery of process taxonomies
  - Outlier detection

**Process Mining** 

## **Part I – Introduction and Basic Concepts**

Context, motivations, goals Characteristics of the analyzed data Classification of Process Mining approaches



Based on slides by Prof. Wil van der Aalst and Dr. Ana Karla A. de Medeiros



## Outline

- Part I Introduction to Process Mining
  - Context, motivation and goal
  - General characteristics of the analyzed processes and logs
  - Classification of Process Mining approaches
- Part II Workflow discovery
  - Induction of basic Control Flow graphs
  - Other techniques ( $\alpha$ -algorithm, Heuristic Miner, Fuzzy mining)
- Part III Beyond control-flow mining
  - Organizational mining
  - Social net discovery
  - Extension of workflow models
- Part IV Evaluation and validation of discovered models
  - Conformance Check
  - Log-based property verification
- Part V Clustering-based Process Mining approaches
  - Discovery of hierarchical workflow models
  - Discovery of process taxonomies
  - Outlier detection



# **Process Mining: basic idea**

- Aims to discover process knowledge based on historical execution data
- Logs register what happened along past process enactments, and are maintained by diverse kinds of transactional IS (WfMS, ERP, CRM,...)





## **Process Mining: basic idea**

The focus is on the real behavior of the process, rather than on its expected/prescribed behavior



# PM vs. Design-time Workflow Analysis

KAR

- Validation bases on comparing models with requirements/expectations
  - Validating real models is hard, and requires some reflection of reality
- Verification concerns the correctness/soundness of the model
  - typically used to answer qualitative questions
    - Is there a deadlock possible?
    - Is it possible to successfully handle a specific case?
    - Will all cases terminate eventually?
    - It is possible to execute two tasks in any order?
- Ex-ante performance analysis
  - Typically regard quantitative aspects
    - How many cases can be handled in 1 hour?
    - What is the average flow time?
  - Common approaches:
    - Simulation, queuing theory
    - Markovian analysis



## **Process Mining vs. Design-time Analysis**

Process mining uses historical event logs as a reflection of reality

- behavioral models are linked to real log events
- Reduces the abstraction gap between model and reality



## **Classification of Process Mining approaches**

# Process Mining can support different kinds of analysis tasks



# Different kinds of knowledge on process execution can be found

#### Control flow perspective:

- What is the typical flow of work for the handling of orders?
- What's the procedure (combination of tasks) followed for orders above 10K?

#### Case perspective:

- Was the invoice 1203 paid on time?
- How regular and rush orders differ in the execution flow ?
- Organizational perspective:
  - Which people appear to be working together closely?



# **Process Mining tasks: Discovery**



### **Discovery: an example control-flow model**



## **Process Mining tasks: Conformance Check**

Compliance Process Model





Auditing/Security





## **Process Mining tasks: Extension**





## **Extension: example of decision point analysis**



## **Process Mining vs Data Mining**



- Process Mining is a specialization of Data Mining
   with a strong business process viewpoint
- Some traditional DM techniques can be used in the context of PM
- New techniques have been specifically developed for process mining
   e.g. the discovery of workflow models



# **Process Mining tools**

#### Open-source tools available at <u>www.processmining.org</u>

- ProM
- ProMimport

🧭 The ProM Framework - Windows Internet Explorer		
COO - Shttp://is.tm.tue.nl/~cgunther/dev/prom/	Google	<b>P</b>
2 A S The ProM Framework	🏠 🔹 🔊 👘 🖶 🔁 <u>P</u> age 🕶 🎯 1	T <u>o</u> ols 👻
	ProM	A Second
	Advanced window management	
tenge de ning te aucour au (untre et au ) te au de la constante de la constan	with the power of more than 190 mining, analysis and conversion plug-ins!	
current version: 4.2 released: 17.9.2007 ProM - the leading proc	ess mining toolkit.	

Processes are an integral part of today's world, driving services and internal functionalities in businesses, governmental bodies, and organizations around the globe. While there are plenty of systems available for supporting the execution of such processes, the current practices for monitoring and analyzing this execution in the organizational reality still leaves a lot to be desired. **Process Mining** is able to fill that can provide conductionant means for





## **ProM architecture**



## ProM



踪 Pro	M [4.2]						_ <del>_</del> _ <del>_</del> ×	
File	Mining Analysis Conversion Exports Window Help							
τ.	Filtered grpd_OneDayConference.xml (Simple filter) ▸	Multi-phase Macro Plu	ıgin					
<u> </u>	Raw grpd_OneDayConference.xml (unfiltered)	Partial Order Generate	pr				ProM	
100	Open new log	Partial Order Aggrega	tor 🛛				and the second se	
	Shou ush ushu	Partial Order Mining T	imeUnit 🔤 🛄 🛛	jrpd_OneDayConference.x	ml		□ □	
		Heuristics miner	a a	rpd_OneDayConfer	ence xml			
		DWS mining plugin	9	ipa_onebayeeniei	chee.xiiii			
		Association Rule Mine	er 📰					
		Genetic algorithm plug	jin 😸	Log filter	Simple Advance			
		Duplicate Tasks GA plug-in	ug-in					
1.5		Alpha algorithm plugin		Process	ses	Start events	Event filter	
		Tsinghua-alpha algorit	thm plugin 🚺 🙍	Only green	n processes will	Only instances starting with	Only green events will be	
		Alpha++ algorithm plu	gin 🦉			a green event will be used.	used.	
		Transition System Ger	nerator		Log	start (complete)	start (complete) GetReady (complete)	
		Region miner		Filter			TravelCar (complete)	
		Language-based Regi	on miner				ConferenceStarts (complet	
		Europe Minor		$\bigcirc$			JoinGuidedTour (complete)	
		Fuzzy Miller		∽∟∟		End events	JoinDinner (complete)	
		Activity Chotosing Min				Only instances ending with	GoHome (complete)	
270		Change mining plugin		Event ty	/pes	a green event will be used.	end (complete)	
		Eromoney abstraction	miner 🖉	Event type	es may be	end (complete)	TravelTrain (complete)	
		Contraction of the second		removed, instances	or their enclosing discarded.		GiveTalk (complete)	
		Social network miner			- 4			
		StanAssignmentminer		comple	ete (keep) 🚽			
-		Organizational miller						
V		Workflow patterns mi	ner <b>e</b>					
	22:01:36 [M] Starting main UI	Case data extraction	plugin				<b>_</b> _	
	22:01:36 [M] ProM Framework Initialized	Logreader Benchmark	(	ti\DDI\D\WS laintDanar_fina	l v4\loaDunninaEvonu	alalarad OneDayConference ye	al using filtor Dofault Log Eiltor	
W	22:03:01 [D] Processing data for buffered log reader initia	LogReader compariso	on driver	(hornowsjointraper_inia	n_vriogranningexam	ne@pu_oneDayconierence.xii	n, using men beraun Log Fine.	
E	22:03:02 [D] NikeFS2: Allocating new swap file. (#1: 67108	Process Instance Insp	pector					
D	22:03:02 [D] NikeFS2: Populating shadow map 1 (of 4 max 22:03:02 [D] Processing data for buffered log reader com	Flower Model Miner						
Т	22:03:02 [D] Processing data for burrered tog reader com 22:03:22 [D] Buffered log reader created from reader buff	k-RI Miner	er					
	22:03:22 [D] Processing data for buffered log reader initia	ated					_	
	22:03:22 [D] Processing data for buffered log reader com	ipleted.						
				1		11		
💐 Sta	rt 📃 🏉 🖂 🍇 🙋 🚀 🛛 🥻 DWSJointPaper_fin 🍕	Vokia PC Suite	G One Touch Access	🜔 http://www.hpovu	3 Microsoft Offic	💌 👯 ProM [4.2] I	T 🛛 < 🧊 🤿 🌉 👘 📢 🕦 22.09	

# Some questions ProM can help answer to

- What an extent the cases (proc. inst.) comply with a process model?
   Where are the problems? How frequent is the (non-)compliance?
- How are the cases actually being executed?
- Statistics on the execution paths of a given model
  - What is the most frequent path?
  - What is the distribution of all cases over the different paths through the process?
  - What are the routing probabilities for each split node?
- Statistics on execution performances
  - What is the average/minimum/maximum throughput time of cases?
  - Which paths take too much time on average? How many cases follow these routings? What are the critical sub-paths for these paths?
  - What is the average service time for each task?
  - How much time was spent between any two tasks in the process model?

## Some questions ProM can help answer to (2)



- Identification and verification of Business rules
  - What are the business rules in the process model?
  - Are the rules indeed being obeyed?
- Interaction among people
  - What is the communication structure and dependencies among people?
  - □ How many transfers happen from one role to another role?
  - Who are important people in the communication flow?
  - Who subcontract work to whom?
  - Who work on the same tasks?





# **Representation of log data: the MXML format**



#### <u>ICAR</u>

## **Event Logs: the MXML format (2)**



## **Event Logs: the MXML format (3)**



Which fields are useful for case-based analyses?

#### **Toy example:** paper reviewing

#### Event log:

pro	cesses
-----	--------

- process instances
  - events

#### Per event:

- activity name
- (event type)
- (originator)
- (timestamp)
- (data)

Google G-	▼ Uitvoere	en o 🧭 🎦 👻	🔂 Bladwi	jzers▼ PageR	ank 🚽 ≫	🔘 Instellir
	(VSS 2007 Pr 6 1 1 1 1 1 1 1 1	avid 🖉 Dul	applic.	🦾 🗸	<b>5</b> - <b>4</b> - <b>5</b>	Page 👻 🚳 Tool
- Timosi	amp >2007-02-2570	0.00.00 000		/Timestomr		
< ninesi < Origina	tors Mike <td>0:00:00.000</td> <td>)+01:00&lt;</td> <td>rimestamp</td> <td>&gt;</td> <td></td>	0:00:00.000	)+01:00<	rimestamp	>	
<td></td> <td></td> <td></td> <td></td> <td></td> <td></td>						
- <audittra< td=""><td>ilEntry&gt;</td><td></td><td></td><td></td><td></td><td></td></audittra<>	ilEntry>					
<td>owModelElement</td> <td>ect <td>wModelEle</td><td>ments</td><td></td><td></td></td>	owModelElement	ect <td>wModelEle</td> <td>ments</td> <td></td> <td></td>	wModelEle	ments		
<event< td=""><td>vpe&gt;complete<td>entType &gt;</td><td></td><td></td><td></td><td></td></td></event<>	vpe>complete <td>entType &gt;</td> <td></td> <td></td> <td></td> <td></td>	entType >				
<timest< td=""><td>amp&gt;2007-03-30T0</td><td>0:00:00.000</td><td>)+01:00&lt;</td><td>/Timestamr</td><td>&gt;</td><td></td></timest<>	amp>2007-03-30T0	0:00:00.000	)+01:00<	/Timestamr	>	
<origina< td=""><td>tor&gt;Mike<td>r&gt;</td><td></td><td></td><td></td><td></td></td></origina<>	tor>Mike <td>r&gt;</td> <td></td> <td></td> <td></td> <td></td>	r>				
<td>ailEntry&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	ailEntry>					
<td>stance&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	stance>					
<ul> <li><processins< li=""> </processins<></li></ul>	tance id="52" descrip	otion="">				
– <audittra< p=""></audittra<>	ilEntry>					
<workfl< td=""><td>owModelElement&gt;<b>inv</b></td><td>ite reviewe</td><td>s</td></workfl<> <td>owModelEle</td> <td>ement&gt;</td> <td></td>	owModelElement> <b>inv</b>	ite reviewe	s	owModelEle	ement>	
<event< td=""><td>Type&gt;<b>start</b><td>/pe&gt;</td><td></td><td></td><td></td><td></td></td></event<>	Type> <b>start</b> <td>/pe&gt;</td> <td></td> <td></td> <td></td> <td></td>	/pe>				
<timest< td=""><td>amp&gt;2006-08-31T0</td><td>0:00:00.000</td><td>)+01:00&lt;</td><td>/Timestamp</td><td>&gt;</td><td></td></timest<>	amp>2006-08-31T0	0:00:00.000	)+01:00<	/Timestamp	>	
<origina< td=""><td>tor&gt;<b>Anne</b><td>)r&gt;</td><td></td><td></td><td></td><td></td></td></origina<>	tor> <b>Anne</b> <td>)r&gt;</td> <td></td> <td></td> <td></td> <td></td>	)r>				
<td>ailEntry&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	ailEntry>					
– <audittra< p=""></audittra<>	ilEntry>					
<workfl< td=""><td>owModelElement&gt;inv</td><td>ite reviewe</td><td>rs</td></workfl<> <td>owModelEle</td> <td>ement&gt;</td> <td></td>	owModelElement>inv	ite reviewe	rs	owModelEle	ement>	
<event< td=""><td>ype&gt;complete<td>entType&gt;</td><td></td><td></td><td></td><td></td></td></event<>	ype>complete <td>entType&gt;</td> <td></td> <td></td> <td></td> <td></td>	entType>				
<1 imest	amp>2006-09-0110	0:00:00.000	)+01:00<	Timestamp	>>	
<origina< td=""><td>-illate c</td><td>or&gt;</td><td></td><td></td><td></td><td></td></origina<>	-illate c	or>				
	allentry>					
- <auditina< td=""><td>iiEritry&gt;</td><td></td><td></td><td></td><td></td><td></td></auditina<>	iiEritry>					
- <data></data>	hute name="recult">	reject / Attr	ibutes			
<td></td> <td>i ejece ( Ata</td> <td>Date</td> <td></td> <td></td> <td></td>		i ejece ( Ata	Date			
<workfl< td=""><td>owModelElement&gt;get</td><td>review 2<!--</td--><td>WorkflowM</td><td>IodelElemer</td><td>nt&gt;</td><td></td></td></workfl<>	owModelElement>get	review 2 </td <td>WorkflowM</td> <td>IodelElemer</td> <td>nt&gt;</td> <td></td>	WorkflowM	IodelElemer	nt>	
<event< td=""><td>Type&gt;complete<td>entType&gt;</td><td></td><td></td><td></td><td></td></td></event<>	Type>complete <td>entType&gt;</td> <td></td> <td></td> <td></td> <td></td>	entType>				
<timest< td=""><td>amp&gt;2006-09-01T0</td><td>0:00:00.000</td><td>)+01:00&lt;</td><td>/Timestamp</td><td>&gt;</td><td></td></timest<>	amp>2006-09-01T0	0:00:00.000	)+01:00<	/Timestamp	>	
<origina< td=""><td>tor&gt;Pete<td>r&gt;</td><td></td><td></td><td></td><td></td></td></origina<>	tor>Pete <td>r&gt;</td> <td></td> <td></td> <td></td> <td></td>	r>				
<td>ailEntry&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	ailEntry>					
– <audittra< p=""></audittra<>	ilEntry>					
- <data></data>						
<attri< td=""><td>bute name="<b>result</b>"&gt;</td><td>reject<td>ibute&gt;</td><td></td><td></td><td></td></td></attri<>	bute name=" <b>result</b> ">	reject <td>ibute&gt;</td> <td></td> <td></td> <td></td>	ibute>			
<td>&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	>					
<workfl< td=""><td>owModelElement&gt;get</td><td>review 1<!--</td--><td>WorkflowM</td><td>lodelElemer</td><td>nt&gt;</td><td></td></td></workfl<>	owModelElement>get	review 1 </td <td>WorkflowM</td> <td>lodelElemer</td> <td>nt&gt;</td> <td></td>	WorkflowM	lodelElemer	nt>	
<event< td=""><td>Type&gt;complete<td>entType&gt;</td><td></td><td><u> </u></td><td></td><td></td></td></event<>	Type>complete <td>entType&gt;</td> <td></td> <td><u> </u></td> <td></td> <td></td>	entType>		<u> </u>		
<timest< td=""><td>amp&gt;2006-09-05T0</td><td>0:00:00.000</td><td>0+01:00&lt;</td><td>Timestamp</td><td>&gt;&gt;</td><td></td></timest<>	amp>2006-09-05T0	0:00:00.000	0+01:00<	Timestamp	>>	
<origina< td=""><td>tor&gt;<b>Pam</b><td>r&gt;</td><td></td><td></td><td></td><td></td></td></origina<>	tor> <b>Pam</b> <td>r&gt;</td> <td></td> <td></td> <td></td> <td></td>	r>				
Audit I r</td <td>allEntry&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	allEntry>					
	ilEntry>	a aut 2 d/W	orleflowMov	del Element	_	
<vvorkti< td=""><td>owwodelelement&gt;tim</td><td>e-out 3<td>OTKTIOWIMO(</td><td>leiElement</td><td>&gt;</td><td></td></td></vvorkti<>	owwodelelement>tim	e-out 3 <td>OTKTIOWIMO(</td> <td>leiElement</td> <td>&gt;</td> <td></td>	OTKTIOWIMO(	leiElement	>	
<=vent	ppe>complete <td></td> <td>+01.00 -</td> <td>Timestam</td> <td></td> <td></td>		+01.00 -	Timestam		
< ninesi <origina< td=""><td>amp 22000-09-1010</td><td>0.00.00.000</td><td></td><td>Timestani</td><td></td><td></td></origina<>	amp 22000-09-1010	0.00.00.000		Timestani		
<td>ailEntry&gt;</td> <td></td> <td></td> <td></td> <td></td> <td></td>	ailEntry>					
- <audittra< td=""><td>ilEntry&gt;</td><td></td><td></td><td></td><td></td><td></td></audittra<>	ilEntry>					
<workfl< td=""><td>owModelElement&gt;col</td><td>lect reviews</td><td><td>wModelEler</td><td>ment&gt;</td><td></td></td></workfl<>	owModelElement>col	lect reviews	<td>wModelEler</td> <td>ment&gt;</td> <td></td>	wModelEler	ment>	
-Event	Super start (EventT)					



# An equivalent (relational) schema for log events



ProM Import allows to convert data from such a database into an MXML file