Business Process Modeling

Kelly Rosa Braghetto

Instituto de Matemática e Estatística
Universidade de São Paulo
kellyrb@ime.usp.br

January 30, 2009

Agenda

1. **Business Process Concepts**
2. **Business Process Modeling**
   - Languages and Formalisms
   - Workflow Patterns
3. **Process Mining**
   - Main Concepts
   - Mining Challenges
   - Related Works
   - Our Goal

Business Process

Business Process

"A set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships".
[Workflow Management Coalition, 1999]

Business Process

Characteristics of a Business Process

- Defined conditions to trigger its initiation
- Defined outputs at its completion
- Variable duration
- Automated activities, capable of workflow management
- Manual activities, which lie outside the scope of workflow management

Process Definition

Process Definition

"The representation of a business process in a form which supports automated manipulation, such as modeling, or enactment by a workflow management system."
[Workflow Management Coalition, 1999]

The process definition consists of a network of activities and their relationships, criteria to indicate the start and termination of the process, and information about the individual activities.

Synonyms: model definition, flow diagram, state transition, flow schematic, workflow script, case type, etc.

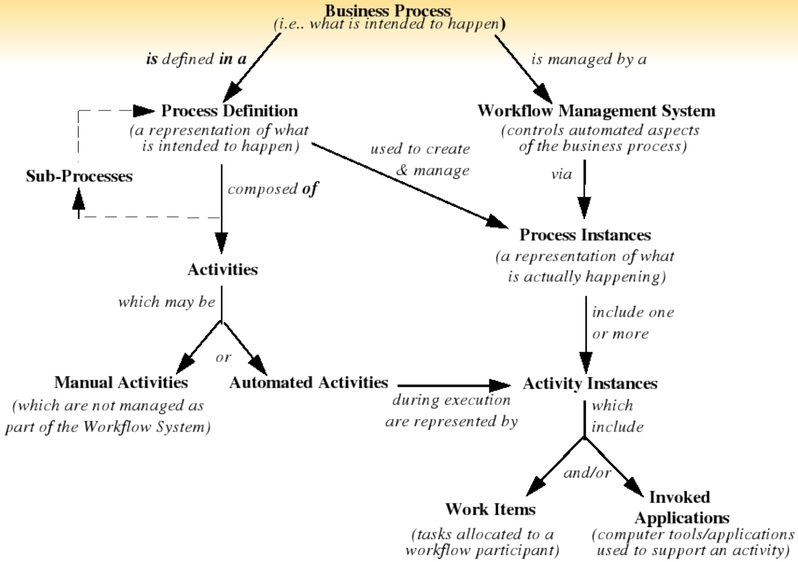
Business Process Management (BPM)

Definition

BPM are methods, techniques, and tools to support the design, enactment, management, and analysis of operational business processes. [Aalst, Hofstede and Weske, 2003]

It can be considered as an extension of classical Workflow Management (WFM) systems and approaches:

- Traditional definitions of workflow systems emphasize the enactment, i.e., the use of software to support the execution of operational processes $\rightarrow$ too restrictive, since they disregard important aspects like diagnosis and simulation
- BPM extends the traditional WFM approach offering support for the analysis phase


Business Process
(i.e.. what is intended to happen)
is defined in a
is managed by a
Process Definition
(a representation of what is intended to happen)
Workflow Management System
(controls automated aspects of the business process)
used to create & manage
Sub-Processes
composed of
via
Process Instances
(a representation of what is actually happening)
Activities
which may be
include one or more
or
Manual Activities
(which are not managed as part of the Workflow System)
Automated Activities
during execution are represented by
Activity Instances
which include
and/or
Work Items
(tasks allocated to a workflow participant)
Invoked Applications
(computer tools/applications used to support an activity)

1 Business Process Concepts

2 **Business Process Modeling**
- Languages and Formalisms
- Workflow Patterns

3 Process Mining

Languages for Process Definition

More suitable to control the execution

- BPMN – Business Process Modeling Notation
- WS-BPEL – Web Service Business Process Execution Language
- XPDL – XML Process Definition Language

Work of general modeling includes arbitrariness and lacks strictness. A diagram modeled with these languages and notations may have various interpretations and one or more different diagrams may denote one process.
→ We must define strict semantics of the models to verify formally them.

Languages for Process Definition

More suitable to be used in analysis

- Automata
  - BPMN $\rightarrow$ automata whose every transition is equipped with a guard (boolean expression)
  - BPEL $\rightarrow$ timed automata
- Petri Nets
  - Workflow Nets (PN extended with cancellation, or-join, multiple instances, etc.)
  - YAWL – Yet Another Workflow Language
- Process Algebras
  - LOTOS + CADP (Construction and Analysis of Distributed Processes) – INRIA VASY
  - π-Calculus, ACP, CCS

Correctness of the Models

Soundness

1. Option to complete – a process, when started, can always complete
2. Proper completion – it should not have any other tasks still running for that process when the process ends
3. No dead transitions – the process should not contain tasks that will never be executed

Workflow Patterns

System Perspectives

- Control-flow perspective – captures aspects related to control-flow dependencies between various tasks (e.g. parallelism, choice, synchronization, etc.)
- Data perspective – deals with passage of information, scoping of variables, etc.
- Resource perspective – deals with resources to task allocation, delegation, etc.
- Exception handling perspective – deals with the exception causes and the actions that must to be taken as result

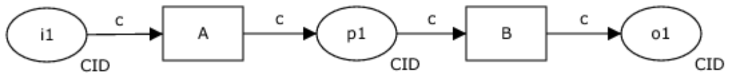
Workflow Patterns

The Workflow Patterns Initiative

- A joint effort of Eindhoven University of Technology (led by Professor Wil van der Aalst) and Queensland University of Technology (led by Professor Arthur ter Hofstede), started in 1999
- Aim: to delineate the fundamental requirements that arise during business process modeling on a recurring basis and describe them in an imperative way
- First important result: set of twenty patterns describing the control-flow perspective of workflow systems, with detailed context conditions and evaluation criteria (2003)

Basic Control-Flow Patterns

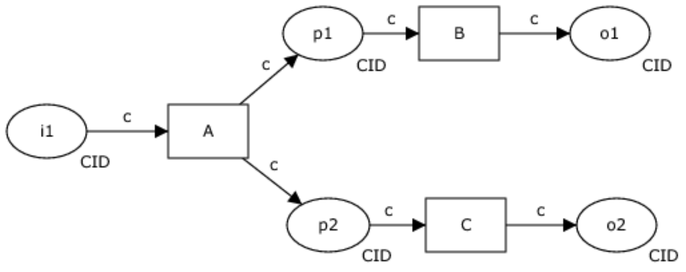
Sequence



$$P = A.B$$

Basic Control-Flow Patterns

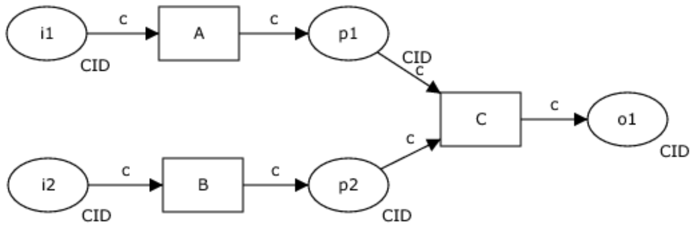
Parallel Split



$$P = A.(B||C)$$

Basic Control-Flow Patterns

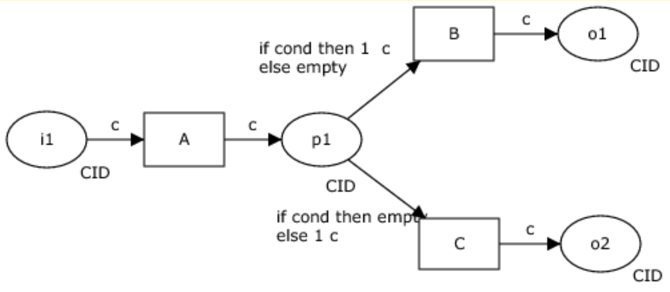
Synchronization



$$P = (A||B).C$$

Basic Control-Flow Patterns

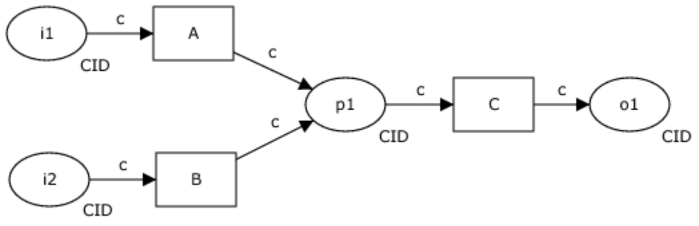
Exclusive Choice



$$P = A.(B + C)$$

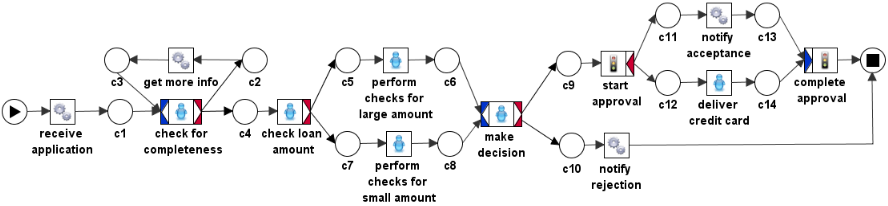
Basic Control-Flow Patterns

Simple Merge



$$P = (A + B).C$$

Example: Applying for a Credit Card



1. Business Process Concepts
2. Business Process Modeling
3. **Process Mining**
   - Main Concepts
   - Mining Challenges
   - Related Works
   - Our Goal

Process Mining

Definition

A process mining algorithm infers an explicit model of a process based on its event logs.

Process Mining is often used when no formal description of the process can be obtained by other means, or when the quality of an existing documentation is questionable.

Process Mining

Analysis

The goal of process mining is to use data logged by the information systems to diagnose the operational processes:

- check conformance between the observed behavior of a system and its modeled behavior, when an “initial” model is available (*Delta Analysis*)
- compare processes
- extract performance measures (i.e., to measure response time, to detect bottlenecks, etc.)

Event Logs

Case ID	Activity	Event Type	Originator	Timestamp	Extra Data
1	File Fine	Completed	Anne	20-07-2004 14:00:00	...
2	File Fine	Completed	Anne	20-07-2004 15:00:00	...
1	Send Bill	Completed	system	20-07-2004 15:05:00	...
2	Send Bill	Completed	system	20-07-2004 15:07:00	...
3	File Fine	Completed	Anne	21-07-2004 10:00:00	...
3	Send Bill	Completed	system	21-07-2004 14:00:00	...
1	Process Payment	Completed	system	24-07-2004 15:05:00	...
1	Close Case	Completed	system	24-07-2004 15:06:00	...
2	Send Reminder	Completed	Mary	20-08-2004 10:00:00	...
3	Send Reminder	Completed	John	21-08-2004 10:00:00	...
2	Process Payment	Completed	system	22-08-2004 09:05:00	...
2	Close case	Completed	system	22-08-2004 09:06:00	...
3	Send Reminder	Completed	John	21-09-2004 10:00:00	...
3	Send Reminder	Completed	John	21-10-2004 10:00:00	...
3	Process Payment	Completed	system	25-10-2004 14:00:00	...
3	Close Case	Completed	system	25-10-2004 14:01:00	...

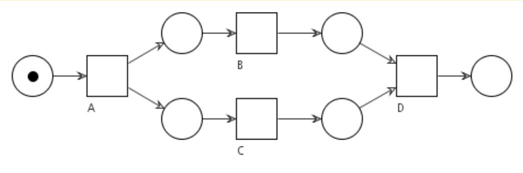
Event Logs

Characteristics:

- Number of activities that a process instance can contain: > 100
- Number of instances that a log can contain: order 10^6
- Size of the space state: $> 2^{100}$ (related to the number of activities of an instance)
- Execution time of an activity: variable

Mining Challenges

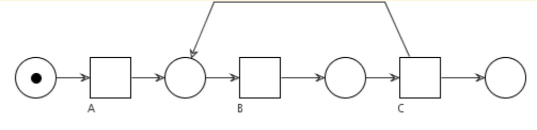
AND-Splits and AND-Joins



$$P \stackrel{def}{=} A.(B||C).D$$

Mining Challenges

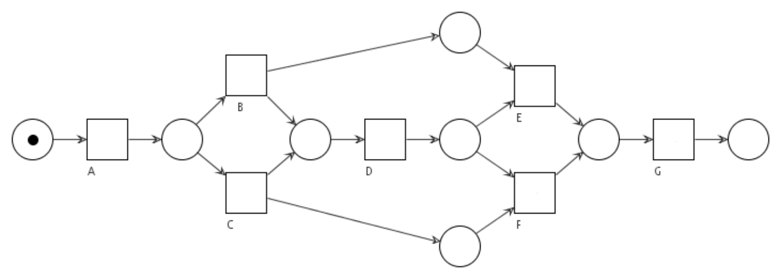
Cycles



$$P \stackrel{def}{=} A.P_1 \qquad P_1 \stackrel{def}{=} B.(C + C.P_1)$$

Mining Challenges

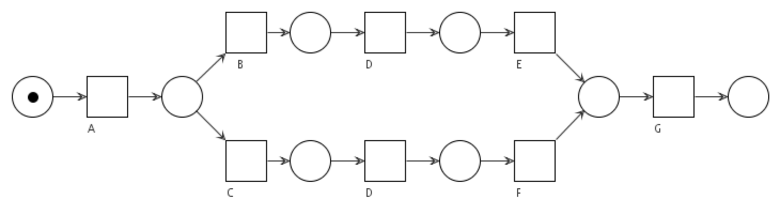
Non-Free-Choices



$$P \stackrel{def}{=} A.((B.D.E) + (C.D.F)).G$$

Mining Challenges

Duplicated Activities



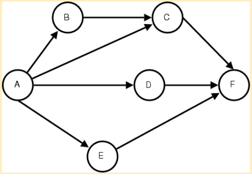
$$P \stackrel{def}{=} A.((B.D.E) + (C.D.F)).G$$

Related Works

Dependencies Graph

- Developed by Agrawal, Gunopulos and Leymann in 1998
- It does not generate an explicit model of the process, but a directed graph representing the dependencies observed between the activities registered in the log
- In this kind of graph, it is not possible to distinguish parallel branches from choice branches in the process structure

$\{(ABCF), (ACDF), (ADEF), (AECF)\}$



Related Works

α-Algorithm and its Extensions

- Developed by Aalst, Weijters and Maruster in 2004
- It generates models in Workflow Nets
- It is based on ordering relations between the activities extracted from the log (ex: direct sequence, direct causality, potential parallelism)
- Using these relations, the algorithm infers the quantity of places of the net and the arcs that connect them to the transitions
- Problems: it does not consider the frequency; it does not deals with non-free-choices and duplicated activities

Related Works

Genetic Process Mining

- Developed by Medeiros, Weijters and Aalst in 2005
- It is able to detect more complex structures (like non-free-choices) due to its genetic approach: it realizes a global search for activities dependencies
- Internal representation: Petri Net (causal matrix)
- Objective function: based on two criteria – completeness and precision
- Genetic operators: include/remove places in the net and modify input/output arcs of the transitions
- Problems: processes with duplicated activities or a large number of activities (huge search space); creation of excessively generalized models

Related Works

MARKOV Method and its Extension

- Developed by Cook and Wolf in 1998
- Uses Markov models to find the sequence of events more probable and algorithmically converts these probabilities in Finite State Machines (FSM)
- Extension for the discovery of concurrency: metrics (entropy, event type counts, periodicity, and causality)
- Problem: the probabilities and the metrics generate a correct model only when they are applied over logs of “well-behaviored” systems. For example: a system that produces events that are directly dependent but that never appear contiguous in the event trace (ex: when a slow thread is mixed in with a fast thread)

Related Works

Limitations

- Complex structures of control-flow (duplicated activities, cycles, parallelisms)
- Processes with a large number of activities
- Noisy data

Other Approaches for Automated Generation of Models

Gramatical Inference

Algorithms for learning stochastic regular grammars or hidden markov models were motivated by the necessity of dealing with noisy scenarios and the unavailability of negative samples.

Examples

- Deterministic Stochastic Finite State Automaton – the algorithm ALERGIA, based on a state merging approach [Carrasco and Oncina, 1994]
- Hidden Markov Models – the Baum-Welch algorithm [Baum, Petrie, Soules and Weiss, 1970], to determine the model parameters

Our Goal

Objective

Extract stochastic models from event logs, to enable performance analysis of business processes.

Candidate formalisms

- Stochastic Automata Networks
- Stochastic Petri Nets
- Stochastic Process Algebras

Thanks for your attention!

References I

- W.M.P. van der Aalst, A. J. M. M. Weijters, and L. Maruster. *Workflow Mining: Discovering Process Models from Event Logs*. IEEE Transactions on Knowledge and Data Engineering, 16(9), pp. 1128–1142, 2004.
- W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, and A. P. Barros. *Workflow Patterns*. Distributed and Parallel Databases, 14(1), pp. 5–51, 2003.
- A.W.M.P. van der Aalst, A. H. M. ter Hofstede, and M. Weske. *Business Process Management: A Survey*. Business Process Management, Lecture Notes in Computer Science, 2678, pp. 1–12, Springer, 2003.

References II

- R. Agrawal, D. Gunopulos, and F. Leymann. *Mining Process Models from Workflow Logs*. In: 6th International Conference on Extending Database Technology, pp. 469–483, 1998.
- L. E. Baum, T. Petrie, G. Soules, and N. Weiss, N. *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*. The Annals of Mathematical Statistics, 41, pp. 164–171, 1970.

References III

- R. C. Carrasco, and J. Oncina *Learning Stochastic Regular Grammars by Means of a State Merging Method* 2nd International Colloquium on Grammatical Inference – ICGI '94, Lecture Notes In Computer Science, 862, pp. 139 – 152, Springer, 1994.
- J. E. Cook, Z. Du, C. Liu, and A. L. Wolf. *Discovering models of behavior for concurrent workflows*. Computers in Industry, 53(3), pp. 297-319, Elsevier Science Publishers B. V., 2004.
- A. K. A. de Medeiros. *Genetic Process Mining* PhD Thesis, Eindhoven University of Technology, 2006.

References IV

- F. Puhlmann, M. Weske. *Using the pi-Calculus for Formalizing Workflow Patterns*. 3rd International Conference in Business Process Management – BPM 2005, Lecture Notes in Computer Science, 3649, pp. 153–168, Springer, 2005.
- WfMC. *Workflow Management Coalition – Terminology & Glossary*. Document Number WFMC-TC-1011, 1999.