Understanding Production Chain Business Process using Process Mining: a Case Study in the Manufacturing Scenario

Barbara Re

Computer Science Division

University of Camerino, Camerino, Italy

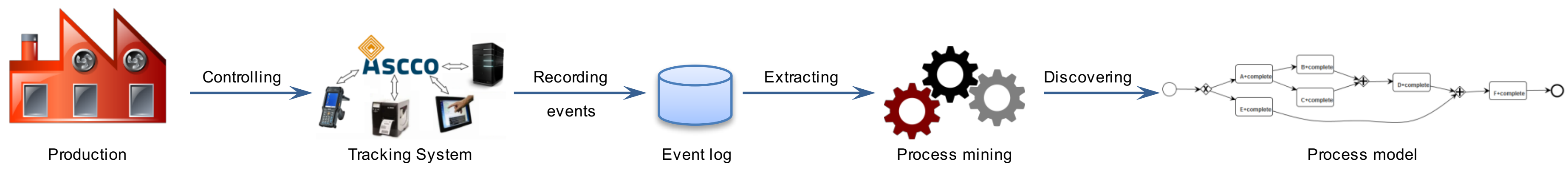
Motivations

- Changes in business context
  - New market conditions
  - Better lead time
- Manufacturing characterized by complex production processes

- Company forced to continuous improvements to advance
- Better understanding of the actual production processes

Process mining provides appropriate techniques

Overall Approach



Process Mining Algorithm

Several Algorithms are available

- α-algorithm
- Fuzzy Miner
- HeuristicsMiner (HM)
- Integer Linear Programming (ILP) Miner
- Inductive Miner (IM)
- Evolutionary Tree Miner (ETM)

Why such algorithms? (I)

- ILP, HM and Genetic Miner have good performance especially with real-life logs[1]
- ETM is the natural progression of the Genetic Miner
- IM is newer than[1], but it outperforms the other three algorithms[2]
- α-algorithm is the reference for the minimum level of performance

[1] De Weerdt, J. et al., 2012. A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, 37(7), pp.654–676.
[2] Leemans, S.J.J., Fahland, D. & van der Aalst, W.M.P., 2014. Discovering block-structured process models from event logs containing infrequent behaviour. In N. Lohmann, M. Song, & P. Wohed, eds. *Lecture Notes in Business Information Processing*. Springer International Publishing, pp. 66–78.

Why such algorithms? (II)

- Algorithms are public available (in ProM 6.5)
- The results may be transformed in BPMN
  - Ensures an unambiguous comparison
  - Only exception is the Fuzzy Miner
    - It returns a fuzzy model that is not a formalism to represent BP
    - Discovered models provide an overall behavior with understandable and high-level information

Case study: a coffee machine company

- Produces professional coffee machines
- The main business is the assembling
  - Only a small portion of the components is made internally
- Production is shared on 6 production lines identified by numbers (1..6)
- Each production line is divided into stations, identified by the letters A to F, with clear tasks
  - Lines from 1 to 4 have six stations
  - Lines 5 and 6 have only five stations
- Production lines, and related stations, are managed by a customized PAIS named ASCCO

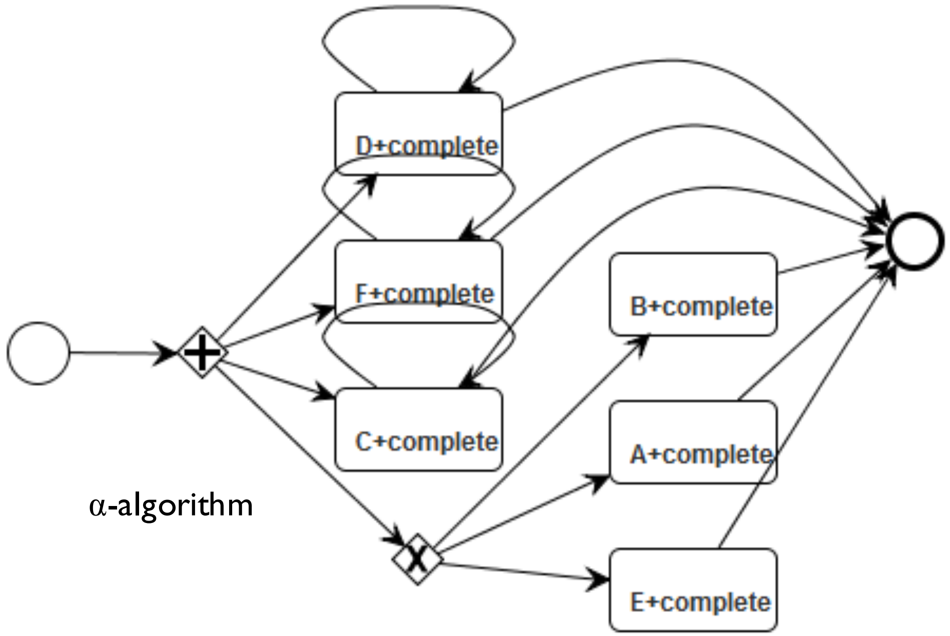
Process mining into practice

- Available more than 450000 event logs
  - 6 years of production
  - related to the manufacturing of 32 different coffee machine
- Logs stored in a relational DB
- Process mining performed with ProM 6.5 framework
- Logs are converted in Extensible Event Stream (XES)[1] format
- Mining algorithm are applied to all 32 sets of log
- Discovered process models are converted to BPMN

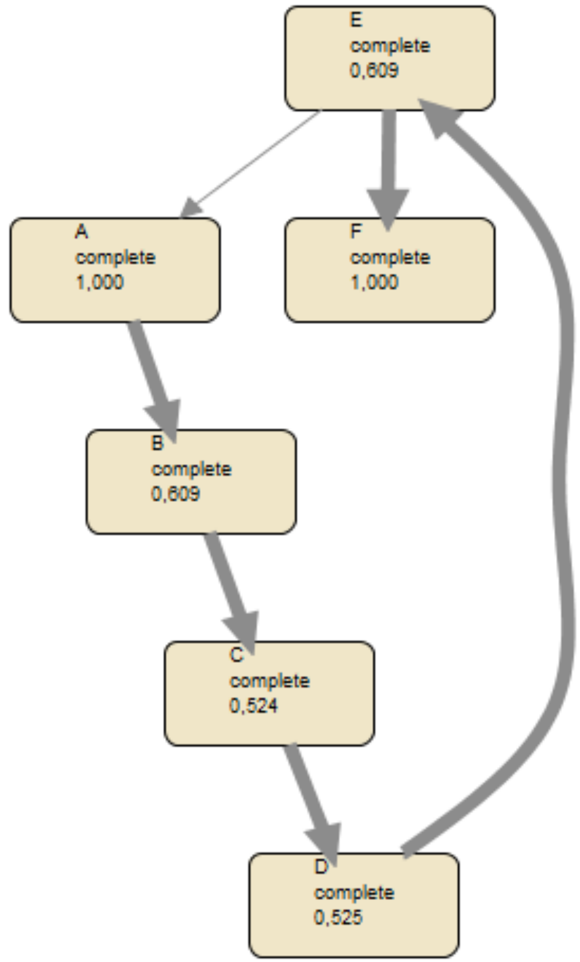
[1] Günther, C.W. & Verbeek, E., 2014. XES Standard Definition ver. 2.0

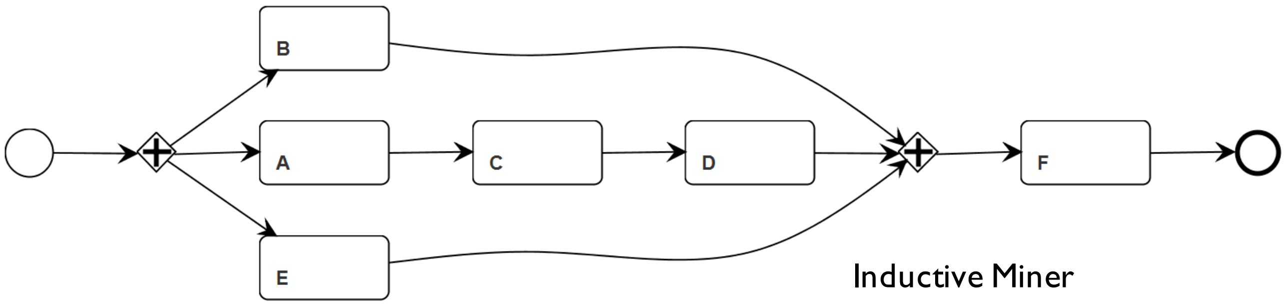
Discovered BP models for coffee machine

Samples for Model 7



α-algorithm





Inductive Miner

Fuzzy model

Some Conclusions about discovered models

- BP models discovered with α-algorithm or ILP are not meaningful for most of the models of coffee machines
  - Confused models
  - Difficult to obtain useful information
- Fuzzy miner is suitable in modelling manufacturing processes
  - Understandable discovered models
  - Activities are highly connected
- HM, IM and ETM return very comprehensible models
- ETM requires several minutes to discover a BP model while other algorithms complete in few seconds

Evaluation criteria for BP models (I)

A BP model may be evaluated according to four quality dimensions:

- Replay fitness expresses the portion of the log behavior that can be replayed by the process model
- Precision is the measure of the level of underfitting
- Generalization is the measure of overfitting
- Simplicity evaluates how easily a human interprets the process model

van der Aalst, W.M.P., 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer Berlin Heidelberg

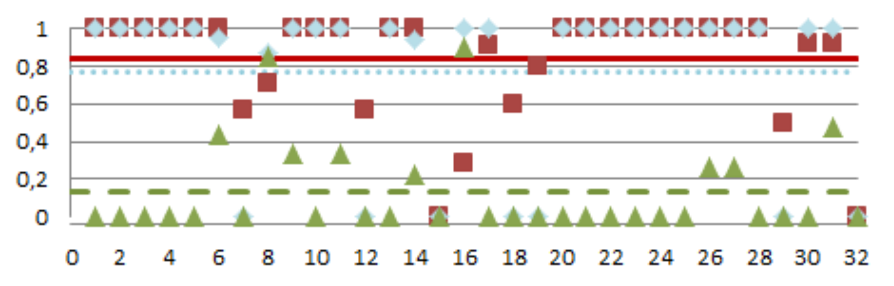
Evaluation criteria for BP models (II)

Complexity metrics for an objective evaluation of Simplicity

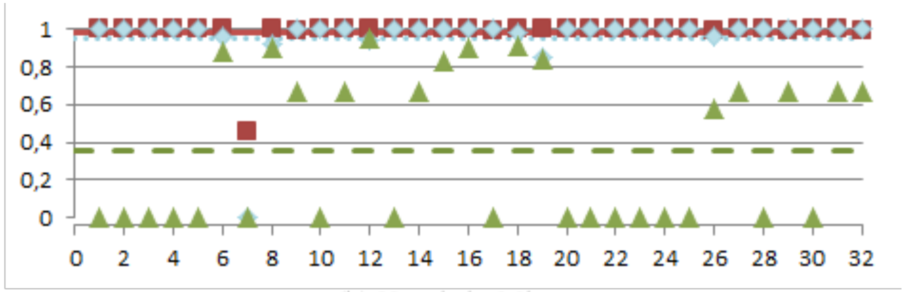
- Size: number of nodes of the model
- Density (Δ): [total number of arcs]/[max number of arcs for the same number of nodes]
- CNC: arcs/nodes
- ACD: number of nodes a connector is in average connected to
- MM: sum of mismatches for each connector type
- CFC: sum over all connectors weighted by their potential combinations of states after a split

Mendling, J., 2008. Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, Springer Publishing Company

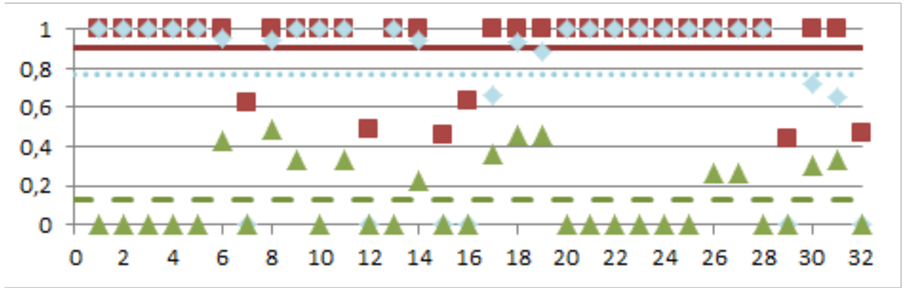
Discovered BP models evaluation (I)



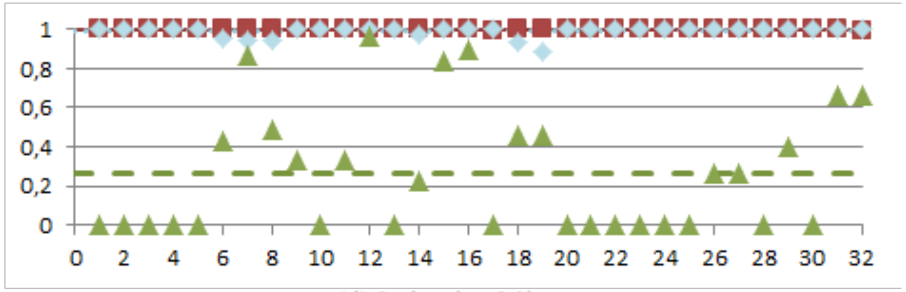
(a) α-algorithm



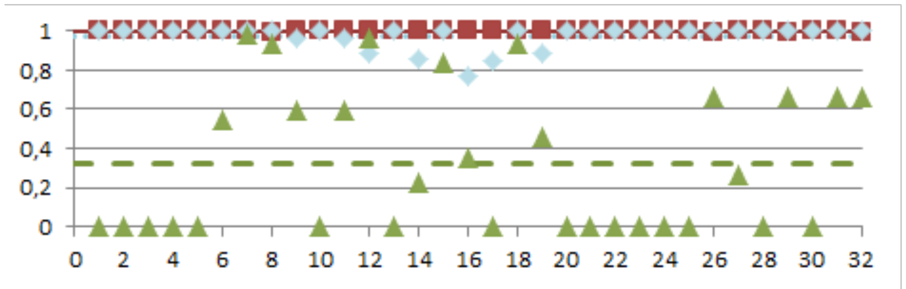
(b) HeuristicsMiner



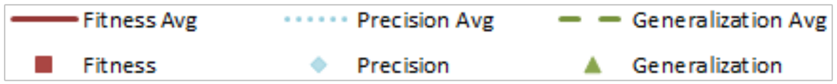
(c) ILP Miner



(d) Inductive Miner



(e) ETM



Discovered BP models evaluation (II)

		Fitness	Precision	Generalization	Size	Δ	CNC	ACD	MM	CFC
7	α	0.5642	0	0	10	0.189	1.7	4.5	7	4
	HM	0.4545	0	0	9	0.139	1.111	3	2	2
	ILP	0.628	0	0	11	0.155	1.545	4.33	6	4
	IM	**0.9998**	0.947	0.866	10	0.122	1.1	4	0	1
	ETM	0.9985	**1**	**0.987**	8	0.125	0.875	0	0	0
8	α	0.7143	0.875	0.854	11	0.109	1.091	3	2	2
	HM	0.9985	0.922	0.901	10	0.122	1.1	3	4	4
	ILP	**1**	0.944	0.493	11	0.109	1.091	3	2	1
	IM	**1**	0.944	0.493	11	0.109	1.091	3	2	1
	ETM	0.9973	**1**	**0.933**	8	0.125	0.875	0	0	0
12	α	0.5644	0	0	9	0.222	1.778	5	4	1
	HM	**0.9988**	0.944	0.956	9	0.125	1	3	2	2
	ILP	0.4851	0	0	13	0.115	1.385	3.4	2	5
	IM	0.9983	**1**	0.964	8	0.125	0.875	0	0	0
	ETM	0.9984	0.885	**0.965**	9	0.125	1	3	2	2
16	α	0.2857	**1**	**0.9**	14	0.093	1.214	3.17	13	5
	HM	0.9995	**1**	**0.9**	8	0.125	0.875	0	0	0
	ILP	0.6323	0	0	11	0.127	1.273	3.33	1	1
	IM	0.9995	**1**	**0.9**	8	0.125	0.875	0	0	0
	ETM	**1**	0.772	0.35	10	0.122	1.1	4	0	1
19	α	0.7966	0	0	10	0.133	1.2	3	2	3
	HM	0.9986	0.848	**0.84**	9	0.139	1.1111	3	4	4
	ILP	**1**	**0.883**	0.458	9	0.139	1.1111	3.5	1	1
	IM	**1**	**0.883**	0.458	9	0.139	1.1111	4	0	1
	ETM	**1**	**0.883**	0.458	10	0.122	1.1	3.33	1	1
29	α	0.4918	0	0	8	0.269	1.875	6	5	1
	HM	0.9968	**1**	**0.667**	7	0.143	0.857	0	0	0
	ILP	0.4384	0	0	8	0.161	1.125	3	2	0
	IM	**0.9996**	**1**	0.4	8	0.143	1	3	2	0
	ETM	0.9968	**1**	**0.667**	7	0.143	0.857	0	0	0

Discovered BP models evaluation (III)

- α-algorithm and ILP have a good performance in a small number of data variations
  - Underperform in all quality dimensions
  - Show significantly high complexity measures
- IM and ETM have best performance on all the logs
  - Highest values of fitness and precision for each log
  - "Good" values of complexity

Discovered BP models evaluation (IV)

- Also HM has good performance
  - High fitness and precision
  - Complexity comparable to IM and ETM
  - The only exception is for log set 7
- Results do not show specific relationship between complexity and quality measures
- IM is preferred over ETM due to its performance

Conclusion and Future Work

- Results have been used for further analysis
- Deviated traces depend on:
  - Run-time errors of PAIS (only a small portion)
  - Incorrect procedures for fixing/replacing faulty components
  - Implementation of some special customizations
- New process awareness drove the company to reorganize production
- Integrating Process Mining in ASCCO

Thanks!!!

Alessandro Bettacchi
alessandro.bettacchi@unicam.it