Brainstorm Seminar Conformance Checking Metrics

Program

8-9 February 2018

An initiative of the scientific research community on process mining

UHASSELT
KNOWLEDGE IN ACTION

KU LEUVEN

GHENT UNIVERSITY

Contents

About

This Brainstorm Seminar on Conformance Checking Metrics is the first seminar organized by the Scientific Research Community on Process Mining. Supported by the Flemish (Belgium) research foundation, this community interchanges research ideas and aspires synergetic research collaborations all over the world. The focus of our research efforts is to close the gap between scientific research on process mining techniques on one hand and the usability of these techniques on the other hand. More specifically, we will address three challenges:

Challenge #1 – Improving usability and understandability for non-experts

Challenge #2 – Understanding and reducing the representational bias in process discovery

Challenge #3 – Improving the metrics to measure the quality of process discovery algorithms

Given these challenges, this seminar aims to harmonize different views on quality metrics for discovery algorithms, and conformance checking in general. The outcome goal is to establish a clear research agenda for conformance checking, which will improve the metrics for quality measurement, as well as the usability and understandability of process discovery for non-experts in the long run.

Organization

The scientific research community on process mining is a Flemish initiative, led by Hasselt University. Ghent University and KULeuven are the other flemish participating partners.
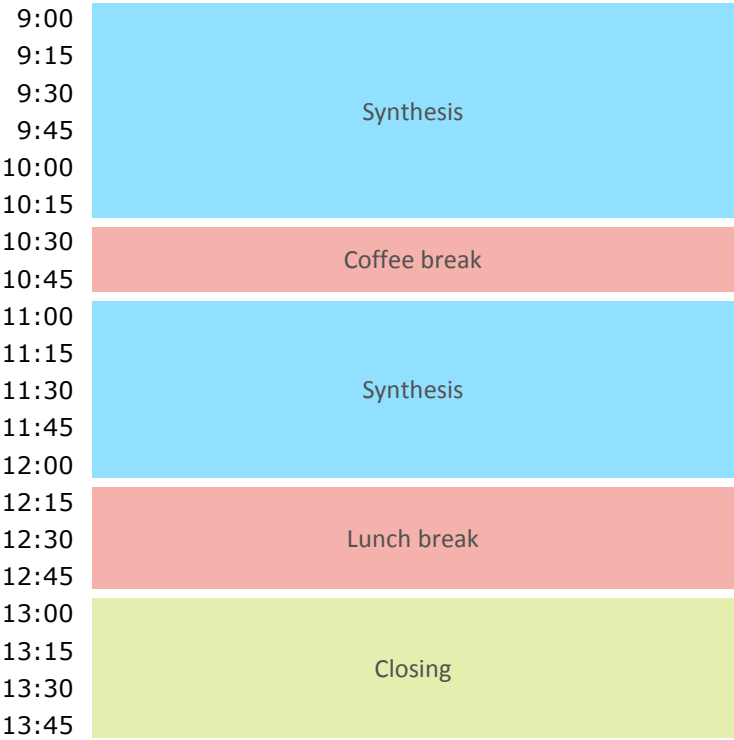
Attendees

Name	Institution
Benoît Depaire	UHasselt
Boudewijn van Dongen	Technische Universiteit Eindhoven
Claudio Di Ciccio	Wirtschafts Universität Wien
Dirk Fahland	Technische Universiteit Eindhoven
Gert Janssenswillen	UHasselt
Jochen De Weerdt	KULeuven
Jorge Munoz-Gama	Pontificia Universidad Católica de Chile (UC)
Josep Carmona	Universitat Politècnica de Catalunya
Koen Vanhoof	UHasselt
Mieke Jans	UHasselt
Seppe vanden Broucke	KULeuven
Wai Lam Jonathan Lee	Pontificia Universidad Católica de Chile (UC)
Wil van der Aalst	RWTH Aachen University

Program Overview

Thursday February 8th, 2018

Time	Session	Introductory Presentation	Chair
9:00	Opening session		
9:15 9:30 9:45	[Discussion] Process mining in practice	Jonathan	Mieke
10:00 10:15	[Discussion] Fitness as primary dimension	Boudewijn	Mieke
10:30 10:45	Coffee break		
11:00 11:15 11:30 11:45	[Discussion] Generalization	Jochen/Seppe	Josep
12:00	[Presentation] Understandability	Benoît	Josep
12:15 12:30 12:45	Lunch break		
13:00 13:15	[Discussion] The need for axioms	Wil	Benoît
13:30 13:45 14:00 14:15 14:30 14:45	[Discussion] Desirable properties (Paradigm independence, confidence, proper measure, …)	Claudio Dirk Gert	Benoît
15:00 15:15 15:30 15:45	Coffee break		
16:00 16:15 16:30	[Discussion] Alignments	Jochen/Seppe Boudewijn	Jorge
16:45	[Presentation] Multi-perspectiveness of metrics	Claudio	Jorge
19:00	Dinner		

Friday February 9th, 2018

Time	Session
9:00 9:15 9:30 9:45 10:00 10:15	Synthesis
10:30 10:45	Coffee break
11:00 11:15 11:30 11:45 12:00	Synthesis
12:15 12:30 12:45	Lunch break
13:00 13:15 13:30 13:45	Closing

Program details

General

Each discussion will be opened by someone giving a short presentation on the concerning topic from his point of view. The idea is to keep these introductions as short as possible (max 15 min., preferable shorter), in order to leave sufficient time for discussion.

Discussions
1	Jonathan
2	Boudewijn
3	Jochen, Seppe
4	Wil
5	Jochen, Seppe, Boudewijn

For discussion 4, Claudio, Dirk and Gert are invited to shortly introduce their requirements/properties for metrics after the general introduction by Wil.

For discussion 5, both Boudewijn and Jochen/Seppe are invited to share their view on the role of alignments in conformance checking.

We will conclude both the morning and afternoon session with a short presentation (15min) of Benoît and Claudio, which will offer further food for thought during the breaks.

Thursday February 8th, 2018

09:15 – 10:00 [Discussion 1] Process mining in practice: the need for use cases

Chair	**Mieke**
Introduction by	**Jonathan**
	In this discussion, the focus lies on the application of conformance checking in practice. What are the reasons for slow adoption by industry? Do we need more use cases? Clearer distinction between *types* of conformance checking? …
Inputs	
Jonathan	"There should be multiple target use cases / criteria for conformance checking such that for each end a different set of output and characteristics applies."
	In recent work, there is a clear research trend on alignment-based approaches. While alignment-based approaches can give detailed conformance diagnostics, such level of detail is not always required. For example, to determine fitness between a log and model, replay-based approaches that make use of the executable property of process models to replay historical traces can be more appropriate when the user only wants to check if the traces can be replayed and would like the conformance results to be more explainable. I envision that conformance checking research could be more focused on the end use where different groups of conformance checking approaches have different objectives. This can perhaps be appreciated from recent research in process discovery where there are multiple research lines that have different objectives. For example, local process models do not focus on start-to-end models but rather frequent patterns within an event log. In contrast, research on inductive miner of the IMD framework aims to handle all the details from a large amount of event data. While the two works focus on very different aspects of process discovery they are both valid and contribute to process discovery. I believe that a potential brainstorm session would end up with a few clear research focuses for conformance checking

	research so that these focuses can be independently pursued but still contribute to the area of conformance checking. For example, one focus could be requiring a conformance checking approach to give deterministic quality measures for determining the quality of discovered models. Another focus could require conformance results to have the property of being able to be projected onto the model for the purpose of diagnosing conformance issues.
Benoît	"We fail to clearly state the purpose of process discovery which prevents proper assessment of the quality of discovered models"
	Discovering/learning models from data can have multiple purposes. One goal is to describe data in a comprehensible way (descriptive analysis). Other goals are to detect remarkable patterns in the data (exploratory), extrapolate outside the sample (inferential), confirm or reject hypothesis (confirmatory), predict future process executions (predictive). Most publications do not explicitly mention the data analysis goal they have in mind. This is troublesome as the way the measure the quality of a model is directly related to its purpose. We should identify proper ways of validating quality dependent on the research goal and motivate the community to cleary state their goals in their research publications.
Jochen Seppe	"Conformance checking in practice and commercial tooling: why is it (close to) non-existent?"
Jorge	"There is a gap between conformance results and conformance understanding."
	Metrics are interesting (e.g., guide evolutionary discovery, ...), but conformance should also pinpoint/sum up where the differences are. This is important for discovery comparison (e.g., two miners could have the same metric with completely different problems detected), but more important for process diagnosis/re-design/... (e.g., what needs to be fixed), but also *Status: Are we there? How understandable are our conformance results? How are the comercial tools doing (e.g., Celonis = long list of violations, ...)* *Thinking lines:* *1- Should we return the deviations as detailed as possible (e.g., 10k misalignments or escaping arcs), sum them up, something in between, or both?* *2- Should we provide the deviations as raw as possible (e.g., to be processed by other algorithms), in terms of common deviation patterns (e.g., how large the patterns, ...), graphically over the model or the log, in natural language, ... 3- Should we provide the deviation where detected, or try to reason a bit (e.g., maybe misalignments on different activities have the same cause, ...)*

10:00 – 10:30 [Discussion 2] Fitness as primary dimension

Chair	**Mieke**
Introduction by	**Boudewijn**
	In this discussion, the focus lies on the dimension (fitness, precision, …) and the connections/hierarchy between them
Inputs	
Boudewijn	"Precision, Generalization and Simplicity are useless without fitting data"
	When discussing the relation between models and logs, we focus on four dimensions. However, fitness is by far the most important one. For log/model combinations with low fitness, precision, generalization and simplicity do not actually mean much. In the past, we have shown that this is the case using generic algorithms. However, it's not the fitness metric that is the issue, it is the actual fitness of the data, i.e. if a model does not fit the data (or the data does not fit a model), measures like precision, generalization and simplicity are useless.

10:30 – 11:00 [Coffee break]

11:00 – 12:00 [Discussion 3] Generalization

Chair	**Josep**
Introduction by	**Jochen** **Seppe**
	How do we tackle generalization?
Inputs	
Jochen Seppe	"Generalization in conformance checking: why are we sticking to odd and bad measures for a concept which is so well understood in the broader data science domain?"

12:00 – 12:15 [Presentation 1] Understandability

Chair	**Josep**
Presentation by	**Benoît**

"What we need is an appropriate understandability measure, not a generalization measure"

Often, one considers four quality metrics for discovered process models: fitness, precision, generalization and complexity/understandability. Compared to the other three, complexity/understandability has received relatively little attention. While this metric is fundamentally different from the other three metrics, as it only considers the model and not the log, we can argue that there are situations where it is more important than generalization. More specifically, if the discovery goal is exploratory rather than confirmatory or inferential, generalization becomes irrelevant and an understandability metric is crucial to rank models as it is always possible to produce a perfectly fitting and precise model.

12:15 – 13:00 [Lunch break]

13:00 – 15:00 [Discussion 4] Axioms or properties? What are desirable properties of metrics?

Chair	**Benoît**
Introduction by	**Wil**
	"The Imprecisions of Precision Measures in Process Mining" Discussion about the need for axioms/properties, and how they should be formalized.
Inputs	**Discussion of additional desirable properties**
	Claudio, Dirk and Gert are invited to introduce the concepts below, after which each of them will be discussed.
Claudio	Paradigm-independence of metrics
	The quality metrics should be paradigm-independent: the nature and measuring of the metrics should not be affected by the modelling language used, let it be imperative, declarative, or mixed. Techniques should not be "applicable by seamless conversions", because unfortunately very often they prove anything but seamless, but rather inherently suitable to any behavioural representation of processes.
Dirk	Quality measures in process mining should be proper metrics or at least proper mathematical measures, on comparable scales.
	We are using quality measures to relate and compare different models to each other. To this end, we aim to "measure" the quality of a model wrt. a particular characteristic in terms of a number, and we compare the numbers with the aim of comparing the characteristic. This inference in the comparison is only valid if the quality that we measure adheres to the mathematical requirements of a "measure". *https://en.wikipedia.org/wiki/Measure_(mathematics)* *When comparing multiple models against their different characteristics, we inevitably also compare/weigh different measures against each other. Thus, the different measures need to be comparable to each other as well. Our intuitive understanding of these measures however goes further as we expect that two models with quality 0.91 and 0.92 are close to each other. For fitness of 0.91 and 0.92, that is more likely true than not (yet it is easy to construct strong counter examples where the models differ on very different parts). For fitness 0.5 and 0.51, the models could have almost nothing to do with each other, yet they seem close. For precision, I cannot even predict how close 0.9 and 0.91 would be.* *A proper metric that also satisfies the triangle inequality would allow to quantify the distance between models more accurately.* *https://en.wikipedia.org/wiki/Metric_(mathematics)*
Gert	Confidence of metrics
	In order to increase the adoption of process mining techniques and to evolve towards a mature research field, it is necessary that the impact and significance of conformance checking results can be measured. Since the results are often directly based on a sample of event data, the question whether they also apply to the real process typically remains unanswered. At this moment, conformance checking does not address concepts such as confidence or uncertainty.

15:00 – 16:00 [Coffee break]

16:00 – 16:45 [Discussion 5] Alignments

Chair	**Jorge**
Introduction by	**Jochen** **Seppe** **Boudewijn**
	The role of alignments in process mining. Jochen, Seppe, Josep and Boudewijn are invited to state their opinions on alignments and their role in conformance checking.
Inputs	
Boudewijn	"Alignments should not be approximated"
	Most metrics in conformance checking these days rely on alignments for their computation. Therefore, approximating alignments leads to errors in the derived metrics which may in turn lead to wrong conclusions about the relation between model and log. Techniques available for approximating alignments (such as decomposition or incremental computations) may hide actual problems in this relation.
Jochen Seppe	"Alignments: is it really worth it?"

16:45 – 17:00 [Presentation 2] Multi-perspectiveness of metrics

Chair	**Jorge**
Presentation by	**Claudio**
	Multi-perspectiveness of metrics
	Is generalization in terms of control-flow the same as generalization in terms of resource allocation? What does it mean to be precise in terms of data flows and associations?

19:00 – xx:xx [Dinner]

Please indicate your attendance and preferences here: https://goo.gl/forms/abwLCKq1LZoqPEhb2

Friday February 9th, 2018

After the discussions on Thursday, the goal on Friday will be to synthesize the different opinions and viewpoints into a clear research agenda for conformance checking.

09:00 – 10:30 [Synthesis]

10:30 – 11:00 [Coffee break]

11:00 – 12:15 [Synthesis]

12:15 – 13:00 [Lunch break]

13:00 – 14:00 [Closing]

Literature

Adriansyah, Arya, Jorge Munoz-Gama, Josep Carmona, Boudewijn F. van Dongen, and Wil M P van der

Aalst. 'Alignment Based Precision Checking'. In Business Process Management Workshops, 137–149. Springer, 2012. http://link.springer.com/chapter/10.1007/978-3-642-36285-9_15.

———. 'Measuring Precision of Modeled Behavior'. Information Systems and E-Business Management, 2015, 1–31.

Broucke, Seppe K L M vanden, Jan De Weerdt, Bart Vanthienen, Jan, and Bart Baesens. 'Determining Process Model Precision and Generalization with Weighted Artificial Negative Events'. Knowledge and Data Engineering, IEEE Transactions On 26, no. 8 (2014): 1877–1889.

Buijs, Joos CAM, Boudewijn F. van Dongen, and Wil M. P. van der Aalst. 'On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery'. In On the Move to Meaningful Internet Systems: OTM 2012, 305–322. Springer, 2012. http://link.springer.com/chapter/10.1007/978-3-642-33606-5_19.

Dongen, B. F. van, J. Carmona, and T. Chatain. 'A Unified Approach for Measuring Precision and Generalization Based on Anti-Alignments', 2016. http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/vDCC-bpm16.pdf.

Janssenswillen, Gert, Niels Donders, Toon Jouck, and Benoît Depaire. 'A Comparative Study of Existing Quality Measures for Process Discovery'. Information Systems 71, no. Supplement C (1 November 2017): 1–15. https://doi.org/10.1016/j.is.2017.06.002.

Janssenswillen, Gert, Toon Jouck, Mathijs Creemers, and Benoît Depaire. 'Measuring the Quality of Models with Respect to the Underlying System: An Empirical Study'. In Business Process Management, edited by Marcello La Rosa, Peter Loos, and Oscar Pastor, 73–89. Lecture Notes in Computer Science 9850. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-45348-4_5.

Munoz-Gama, Jorge. Conformance Checking and Diagnosis in Process Mining: Comparing Observed and Modeled Processes. Vol. 270. Springer, 2017. https://books.google.be/books?hl=en&lr=&id=cLaVDQAAQBAJ&oi=fnd&pg=PR7&dq=conformance+checking+and+diagnosis+in+prQJcaQ.

Muñoz-Gama, Jorge, and Josep Carmona. 'A Fresh Look at Precision in Process Conformance'. In Business Process Management, 6336:211–226. Hoboken, NJ, USA: Springer, 2010. http://link.springer.com/chapter/10.1007/978-3-642-15618-2_16.

Rogge-Solti, Andreas, Arik Senderovich, Matthias Weidlich, Jan Mendling, and Avigdor Gal. 'In Log and Model We Trust? A Generalized Conformance Checking Framework'. In Business Process Management, edited by Marcello La Rosa, Peter Loos, and Oscar Pastor, 179–96. Lecture Notes in Computer Science 9850. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-45348-4_11.

Tax, Niek, Xixi Lu, Natalia Sidorova, Dirk Fahland, and Wil MP van der Aalst. 'The Imprecisions of Precision Measures in Process Mining'. ArXiv Preprint ArXiv:1705.03303, 2017. https://arxiv.org/abs/1705.03303.

Practical Information

FRC The Field Research Centre - UHasselt is located on <u>Terhills</u> in the periphery of <u>Connecterra</u>, the main gate of the only national park in Belgium, the National Park Hoge Kempen.

To reach the FRC, take exit 33 on the E314 highway (Brussels - Aachen) and follow the signs 'Maasmechelen Leisure Valley' on a route that follows the canal for some kilometers. Keep following this road across tree roundabouts until the signs 'PARKING EUROSCOOP' and 'CONNECTERRA' indicate the parking lot. Continue the rest of the way (300m) walking through the French Garden as indicated below.

(1) Field Research Center
(2) Hotel & Restaurant Terhills
(3) Parking lot

Location
Field Research Centre
Zetellaan 52
Maasmechelen, Belgium

Hotel **Hotel Terhills**
Zetellaan 68
Maasmechelen, Belgium

Note Attendants are allowed free entrance to the Connectera Nature Park on both days.