Exploiting Process Cubes, Analytic Workflows and Process Mining for Business Process Reporting: A Case Study in Education

Alfredo Bolt[1], Massimiliano de Leoni[1], Wil M. P. van der Aalst[1], and Pierre Gorissen[2]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
{a.bolt,m.d.leoni,w.m.p.v.d.aalst}@tue.nl
[2] Hogeschool van Arnhem en Nijmegen, Nijmegen, The Netherlands
pierre.gorissen@han.nl

Summary. Business Process Intelligence (BPI) is an emerging topic that has gained popularity in the last decade. It is driven by the need for analysis techniques that allow businesses to understand and improve their processes. One of the most common applications of BPI is *reporting*, which consists on the structured generation of information (i.e., reports) from raw data. In this article, state-of-the-art process mining techniques are used to periodically produce automated reports that relate the actual performance of students of Eindhoven University of Technology to their studying behavior. To avoid the tedious manual repetition of the same process mining procedure for each course, we have designed a *workflow* calling various process mining techniques using RapidProM. To ensure that the actual students' behavior is related to their actual performance (i.e., grades for courses), our analytic workflows approach leverages on *process cubes*, which enable the dataset to be *sliced* and *diced* based on courses and grades. The article discusses how the approach has been operationalized and what is the structure and concrete results of the reports that have been automatically generated. The reports were sent to lecturers and feedback was collected through an evaluation form. Also, we analyzed an example report to show the range of insights that they provide.

Key words: Business Process Reporting, Analytic Workflows, Process Mining, Process Cubes, Education.

1 Introduction

Business Process Reporting (BPR) refers to the provision of structured information about processes in a regular basis, and its purpose is to support decision makers. Reports can be used to analyze and compare processes from many perspectives (e.g., performance, costs, time). In education, for example, it is interesting to study student's grades in a course over different years, and how these grades are related to the students' behavior. In order to be effective, BPR presents some challenges:

1. It should provide insights about the typical behavioral characteristics (e.g., throughput time or resource utilization) and highlights the issues (e.g., bottlenecks).

2. It should be repeatable (i.e., not require great efforts to repeat the analysis).
3. It should be able to drill down into the process data and compare the different groups and process variants to highlight dissimilarities.

This paper discusses how to address the three challenges mentioned above by combining *process mining*, *analytic workflows* and *process cubes*.

Process mining is a relatively young research discipline that is concerned with discovering, monitoring and improving real processes by extracting knowledge from event logs readily available in today's systems [1]. This allows the extraction of insights about the overall and inner behavior contained in any given process. Hundreds of different process mining techniques have been proposed in literature. These are not limited to process-model discovery and the checking of conformance. Also, other perspectives (e.g., data) and operational support (e.g., predictions) are included. Process mining is supported by commercial and academic software tools, such as Disco and ProM [2].[1]

When dozens of different reports need to be produced, it can be tedious and error-prone to repeat all the process-mining analyses to be incorporated in the reports. Process mining tools such as ProM are not designed to automatically repeat the application of the same process-mining analyses on an arbitrary number of (sub sets of) event logs. Therefore, it is not possible to automatically generate any arbitrary number of reports. Here we combine process mining with analytic workflow systems, which allow one to design, compose, execute, archive and share workflows that represent some type of analysis or experiment. Each activity/step of an analytic workflow is one of the steps to conduct a non-trivial process-mining analysis, which can ange from data filtering and transformation to process discovery or conformance checking. Once an analytic workflow is configured, it can be executed as many times as needed without the reconfiguration. Analytic workflows are specialization of Scientific workflows tailored towards analytic purposes. Scientific workflows have successfully been applied in many settings [3, 4]. Bolt et al. [5] illustrate the formalization and operationalization of a framework to support process-mining analytic workflows where the steps are linked to the application of process-mining techniques.

Business Process Reports are usually intended to provide a comparative analysis of the differences observed in the different variants of process executions (e.g., executions for gold versus silver customers). Therefore, the data needs to be split into sub sets, where the same analysis needs to be repeated for each set (e.g., the process discovery) and the multiple results need to be compared (e.g., the discovered models need to be compared for differences). Process cubes [6, 7] are used to overcome this issue: in a process cube, events are organized into cells using different dimensions. The idea is related to the well-known notion of OLAP (Online Analytical Processing) data cubes and the associated operations, such as slice, dice, roll-up, and drill-down. By applying the correct operations, each cell of the cube contains a sub-set of the event log that complies with the homogeneity assumption mentioned above.

This paper shows through a case study how these three ingredients (process mining, process cubes and analytic workflows) can be mixed for business process reporting. The case study presented in this paper is about a business-process reporting service at Eindhoven University of Technology (TU/e). The service produces a report each year

[1] ProM tools is free to download from `http://www.promtools.org`

for each course that is provided with video lectures. The report is sent to the responsible lecture and provides insights about the relations between the use of students of the video lectures and their final grades on the course. The usefulness of the reports are evaluated with dozens of lecturers. While the involved lecturers suggested a number of improvements, they also clearly acknowledged the added value of those reports to understand the usefulness of watching video lectures.

The remainder of this paper is organized as follows. Section 2 provides an overview of the case study and discusses the structure of the desired reports. Sections 3 and 4 summarize the main concepts related to process cubes and analytic workflows and illustrates how they are concretely applied in this case study. Section 5 illustrates the resulting report for one of the courses as well as it discusses the results of the evaluation with the lecturers. Finally, Section 6 concludes the paper.

2 A Case Study in Education

Eindhoven University of Technology (TU/e) provides video lectures for many courses to support students who are unable to attend face-to-face lectures for various reasons. Student usage of video lectures and their course grades are logged by the TU/e systems. The purpose of this case study is to show how raw data extracted from TU/e systems can be transformed into reports that show insights about students' video lecture usage and its relation with course grades by using process mining, process cubes and analytic workflows. Figure 1 describes the overview of this case study.

The data used in this case study contains **246.526** *video lecture views* and **110.056** *course grades* of **8.122** students, **8.437** video lectures and **1.750** courses at TU/e for the academic year 2014-2015. Each student and course has a unique identifier code (i.e., *student id, course code*). The data reveals enormous variability; e.g., thousands of students watch video lectures for thousands of courses and every course has a different set of video lectures, and they have different cultural and study backgrounds, which leads to different behavior. Therefore, we need to provide different reports and, within a report, we need to perform a comparative analysis of the students when varying the grade.

Before describing our approach and the ingredients used, we sketch the report we aim for. The report is composed of three sections: *course information*, *core statistics* and *advanced analytics*, as shown in Figure 1.[2] The analysis results refer to all students who registered for the course exam, independently whether or not they participated in it.

The course information section provides general information, such as the course name, the academic year, the number of students, etc. The core statistics section provides aggregate information about the students, such as their gender, nationality, enrolled bachelor or master program, along with course grades distribution and video

[2] An example report, where student information has been anonymized, can be downloaded from `http://www.win.tue.nl/~abolt/userfiles/downloads/Reports/sample.zip`

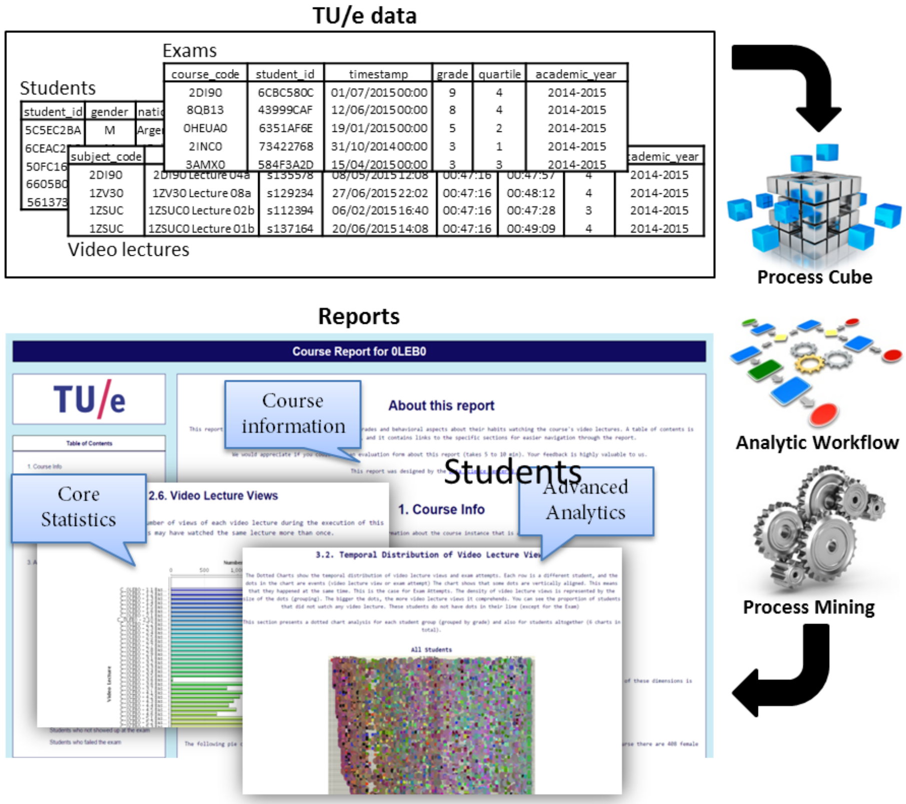


Fig. 1: Overview of the case study: TU/e data is transformed into reports by using process mining, process cubes and analytic workflows.

lecture views. The advanced analytics section contains more detailed diagnostics obtained through process mining techniques. It is divided in three parts and leverages on process mining techniques:

- The level of compliance wrt. the *natural* way of watching the video lectures, which is watching all video lectures sequentially before participating in the exam. This is achieved by integrating the process cube and the conformance-checking technique presented in [8]. An example result is shown in Figure 2a.
- Dotted charts [9] highlighting the temporal distribution of video-lecture watching. Each row indicates a different student and the X axis is the time dimension. Each row contains a dot for each event referring to the respective student. The dot are distributed along the X axis according to the time when the event occurred. An example chart is shown in Figure 2b.
- Process models that show the most frequent patterns and deviations of watching video lectures by students. An example of a model that shows the most frequent

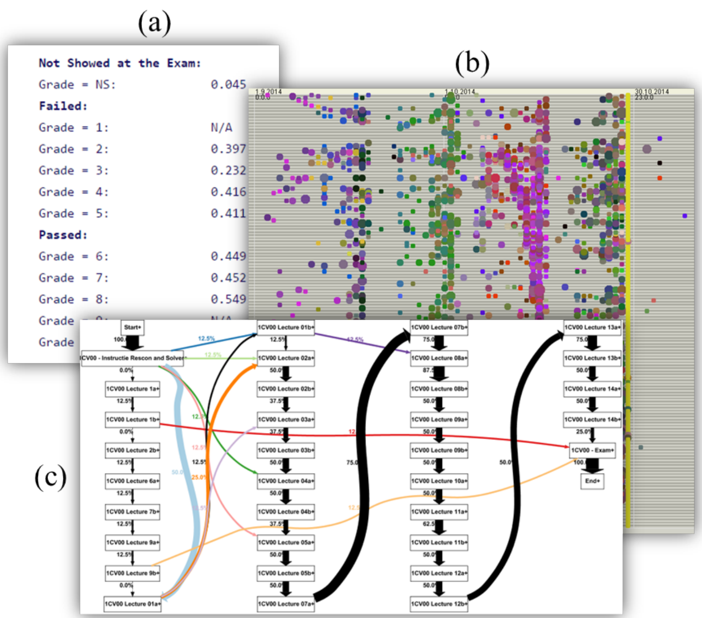


Fig. 2: Examples of results obtained using process mining techniques: (a) Conformance Checker, (b) Dotted Chart, (c) Sequence analysis.

order followed by students to watch video lectures is shown in Figure 2c. In these models, edge thickness represents relative frequency. The edges that correspond to the "natural" order are colored black. From the edges that do not correspond to the "natural" order, only 10 edges are shown: those with the highest relative frequency. In order to avoid confusion by the overlapping of the edges, we used different colors for each of these edges.

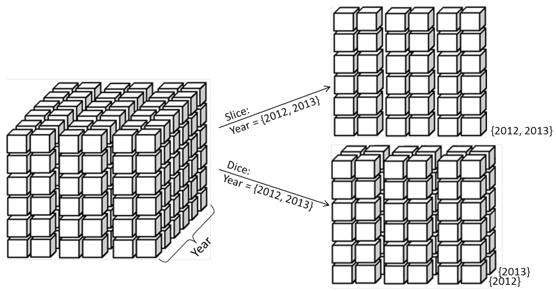
The next two sections show how the desired reports can be generated using our tools, followed by an evaluation of the approach.

3 Process Cubes as a Means to Select and Compare

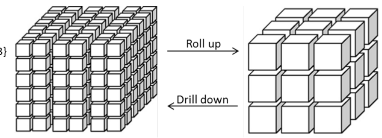
This section discusses the basic concepts of process cubes and illustrates how they have been applied in the generation of the reports.

3.1 Basic Concepts

Processes are not static within moderns organizations but their instances continuously adapt to the dynamic context requirements of modern organizations. Therefore, an event



(a) Slice and Dice



(b) Roll up and Drill down

Fig. 3: Schematic examples of cube operations

log records executions of several process variants, whose behavior depends on the context. As a consequence, the event log needs to be split to obtaining sub-logs, each of which records the events pertaining a certain variant. This is commonly handled by manually filtering the event data. However, this approach is unpractical in scenarios if many different process variations exist.

The *process cubes* approach [10, 6] provides the capabilities needed to perform multidimensional analysis over event data so it can be split and events can be grouped according to their data values. This allows one to isolate and analyze the different process variations of a process given the available data.

Process cubes are inspired by the idea of structuring and scoping the event data through classical OLAP operations (i.e., slice, dice, roll up, drill down). However, because we are dealing with events we cannot use standard OLAP tools. See [10, 6] for details.

A process cube is characterized by a set of *dimensions*, each of which is associated with one or a group of event's data properties. For each combination of values for the different dimensions, a cell exists in the process cube. Hence, each process-cube cell contains the events that assign certain values to the data properties. Each cell of the process cube contains event data that can be used by process mining techniques as illustrated by Figure 2. Please note that certain dimensions may be considered as irrelevant and, therefore, they are ignored and are not *visible* in the cube. Also, some dimensions may be not readily available in the event data; however, they can be derived from the existing dimensions. For example, the "Year" and "Day" dimensions can be derived from the "Timestamp" dimension.

The *slice* operation selects a subset of values of a dimension while removing that dimension from the analysis. For example, if the "Year" dimension is sliced for Year = {2012, 2013}, only the events in those years are retained. Also, the "Year" dimension is removed from the cube as shown in Figure 3a. The latter implies that cells with different values for the "Year" dimension and the same values for the other dimensions are merged.

The *dice* operation is similar to the *slice* operation, with the difference that the dicing dimension is retained. So, the dice operation is only removing cells without merging any cells: the dicing dimension can still be used for further exploration of the event data, as shown in Figure 3a.

The *roll up* and *drill down* operations change the granularity level of a dimension. As shown in Figure 3b, if a dimension is rolled up, an attribute with a more coarse-grained granularity will be used to create the cells of the cube, and if a dimension is drilled down, an attribute with a more fine-grained granularity will be conversely used. For example, the “Day” dimension can be rolled up to “Month”, and the “Month” dimension can be drilled down to “Day”.

3.2 Application to the Case Study

For performing process cube operations over the TU/e data we used the *Process Mining Cube* (PMC) tool introduced in [6]. As mentioned before, the starting point is an event data set. This set has been obtained by defining and running opportune joins of tables of the database underlying the video-lecture system of TU/e (see Section 2). A fragment of the event data set is shown in Table 1.

Table 1: A fragment of event data generated from TU/e system: each row corresponds to an event.

Event Id	Student Id	Nationality	Education Code	Course Code	Activity	Quartile	Academic Year	Timestamp	Course Grade	...
1	1025	Dutch	BIS	2II05	Lecture 1	1	2014-2015	03/09/2012 12:05	6	...
2	1025	Dutch	BIS	2II05	Lecture 2	1	2014-2015	10/09/2012 23:15	6	...
3	1025	Dutch	BIS	1CV00	Lecture 10	3	2014-2015	02/03/2012 15:36	7	...
4	2220	Spanish	INF	1CV00	Lecture 1	3	2014-2015	20/03/2013 16:24	8	...
5	1025	Dutch	BIS	2II05	Exam	2	2014-2015	13/12/2012 12:00	6	...
6	2220	Spanish	INF	1CV00	Lecture 4	3	2014-2015	25/03/2013 11:12	8	...
7	2220	Spanish	INF	1CV00	Exam	3	2014-2015	04/04/2013 12:00	8	...
...	...	...	...	...	...	...	...	...	...	...

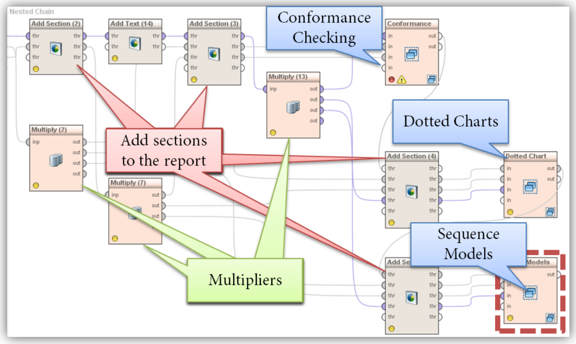
Using the event data, we created a process cube with the following dimensions: Student Id, Student Gender, Student Nationality, Student Education Code, Student Education Phase, Course Code, Course Department, Activity, Activity Type, Grade, Timestamp, Quartile and Academic Year.

After *slicing* and *dicing* the cube, it contained **87.500** cells: one for each combination of values of the “Course Code”, “Quartile” and “Course Grade” dimensions. Each cell corresponds to an event log that can be analyzed using process mining techniques.

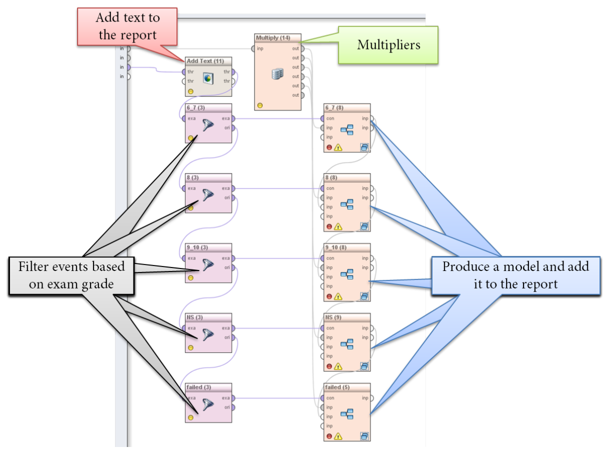
4 Analytic Workflows as a Means to Automate Analysis

Process mining experiments usually require analysts to perform many analysis steps in a specific order. As mentioned in Section 1, it is not unusual that the same experiment has to be carried out multiple times. This is normally handled by manually executing the analysis steps of the experiment, thus requiring large periods of time and resources and introducing the risk of human-induced errors.

Analytic workflows can be used to address this problem. They are defined by chaining analysis and data-processing steps, each of which consumes input produced by previous steps and generates output for the next steps. The analytic workflows can be



(a) Advanced Analytics section sub-workflow



(b) Explosion of the "Sequence Models" sub-workflow

Fig. 4: Implemented analytic workflow used to generate the reports. Each instance of a course can be autimatically analyzed in this way resulting in the report described.

applied to any situation where the same analysis needs to be repeated again and again. The application of analytic workflows to process mining analysis is discussed in [5].

For automatically generating the reports we used RapidProM [11, 5], which extends the RapidMiner analytic workflow tool with process mining techniques.[3]

Figure 4a illustrates the analytic workflow that is used to generate each report. Figure 4b shows the explosion of the "Sequence Models" section of the analytic workflow.

The operators shown in Figure 4 are used for different purposes: *Multipliers* allow one to use the output of an operator as input for many operators. *Filter* operators select a subset of events based on defined criteria. *Process mining operators* are used to produce

[3] Free version and installation instructions for RapidMiner and the RapidProM extension are available at `http://www.rapidprom.org` or at the RapidMiner *Marketplace*.

analysis results. For example, the operators highlighted in blue in Figure 4b produce a sequence model from each filtered event data.

The complete RapidProM implementation of the analytic workflow used in this case study is available at `http://www.win.tue.nl/~abolt/userfiles/downloads/Reports/report.rmp`. Readers can execute this workflow in RapidMiner to generate a report using the sample event log available at `http://www.win.tue.nl/~abolt/userfiles/downloads/Reports/sample.xes`.[4]

5 Evaluation

We applied our approach that combines analytic workflows and process cubes to the case study presented in Section 2. Concretely, we generated **8.750** course reports for 1750 courses given at TU/e in each of the 5 quartiles (i.e., 4 normal quartiles + *interim* quartile) of the academic year 2014-2015. For reliability of our analysis, we only considered those courses where, on average, each student watched at least 3 video lectures. In total, 89 courses were selected. For each one of these courses, an automatically generated report was sent to the corresponding lecturer.

Section 5.1 provides a detailed analysis of the findings that we could extract from the report for a particular course. Along with the report, we also sent an evaluation form to the lecturers. The purpose of the evaluation forms is to verify whether lecturers were able to correctly interpret the analysis contained in the report. The results are discussed in Section 5.2.

5.1 Report For an Example Course

To illustrate the contents and value of the report, we selected an example course: "Introduction to modeling - from problems to numbers and back" given in the third quartile of the academic year 2014-2015 by the Innovation Sciences department at TU/e. This course is compulsory for all first-year students from all programs at TU/e. In total, 1621 students attended this course in the period considered. This course is developed in a "flipped classroom" setting, where students watch online lectures containing the course topics and related contents, and in the classroom, they engage these topics in practical settings with the guidance of the instructor.

The video lectures provided for this course are mapped onto weeks (1 to 7). Within each week, video lectures are numbered to indicate the order in which students should watch them (i.e., 1.1 correspond to the first video lecture of the first week). As indicated by the course's lecturer, the first video lectures of each week contain the course topics for that week, and the last video lectures of each week contain complementary material (e.g., workshops, tutorials). The number of video lectures provided for each week depends on the week's topics and related activities, hence, it varies.

[4] When running the workflow, make sure that the *Read File* operator points to the sample event log and the "HTML output directory" parameter of the *Generate Report* operator points to the desired output folder.

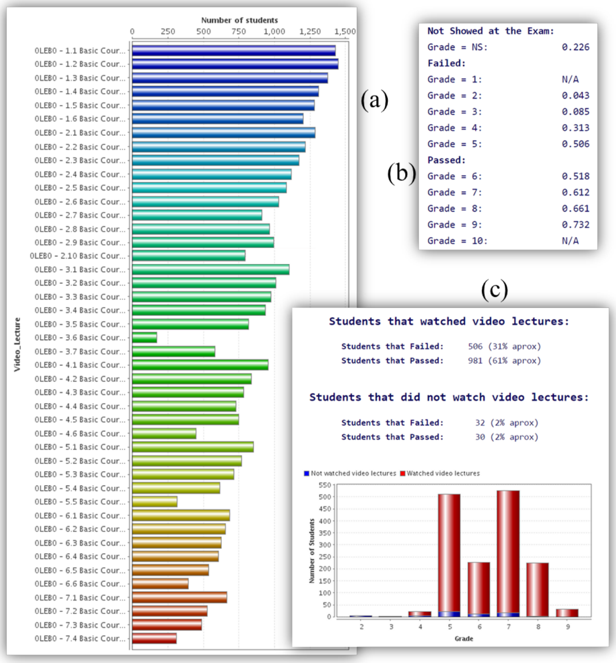
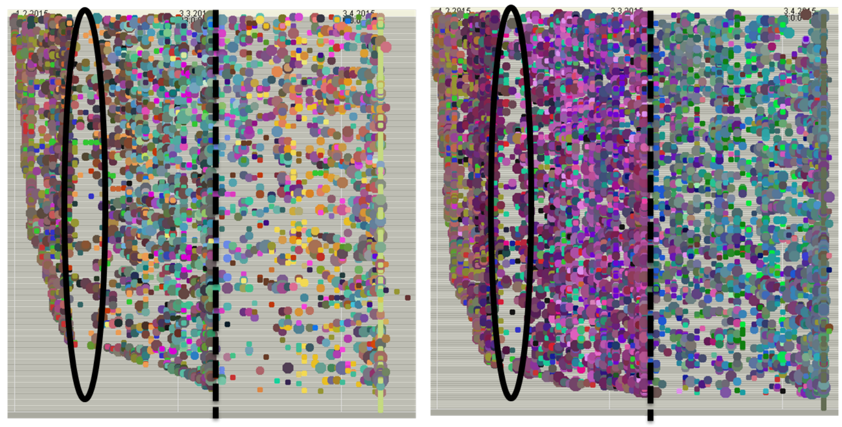


Fig. 5: Analysis results contained in the report of the course 0LEB0:
(a) Number of students that watched each video lecture
(b) Compliance with the natural viewing order by course grade
(c) Grades distribution for students who watched video lectures (in red) or did not (in blue)

Figure 5.a shows for each video lecture the number of students that watched it. We can observe that the number of students that watch the video lectures decreases as the course develops: most students watched the video lectures corresponding to the first week (i.e., 1.X) but less than half of them watched the video lectures corresponding to the last week (i.e., 7.X). Note that within each week, students tend to watch the first video lectures (i.e., X.1, X.2) more than the last ones (i.e., X.5, X.6). This was discussed with the course's lecturer. It is explained by the fact that, as mentioned before, the first video lectures of each week contain the topics, and the last ones contain complementary material.

Figure 5.b shows for each student group (i.e., grouped by their grade) the level of compliance, averaged over all students in that group, of the real order in which students watch video lectures, compared with the "natural" or logical order, namely with watching them in sequence (i.e., from 1.1 to 7.4). The compliance level of each student is measured on a scale from 0 to 1, where 1 indicates that the student has watched all video lectures in the logical order and 0 indicates that no video lectures were viewed in



(a) Grade = 5 (failed) (b) Grade = 6 or 7 (passed)

Fig. 6: Dotted charts for students grouped by their course grades

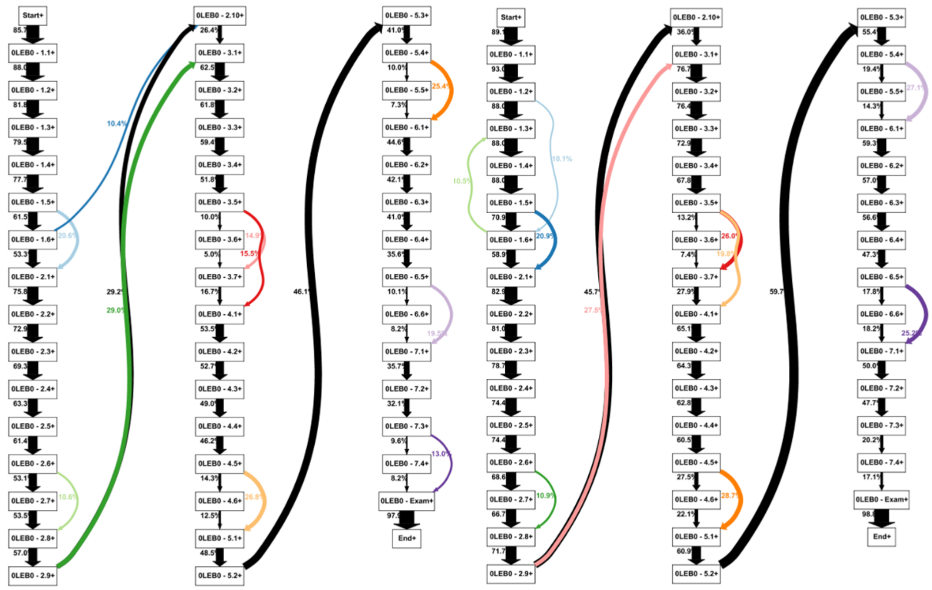
the logical order or, even, not watched at all. We can observe that students with higher grades have higher levels of compliance than students with lower grades.

Figure 5.c shows the grade distribution for this course where each bar is composed by two parts corresponding to the number of students who watched at least one (red part) video lecture and the number of students who did not (blue part). We can observe that the best students (i.e., with a grade of 8 or above) use video lectures. On the other hand, we observe that watching video lectures does not guarantee that the student will pass the course, as shown in the columns of students that failed the course (i.e. grade ≤ 5).

Figure 6 shows dotted chart analysis described in Section 2 for two student groups: (a) students that failed the course with a grade of 5, and (b) students that passed the course with a grade of 6 or 7. Each row corresponds to a student and each dot in a row represents that student watching a video lecture or taking the exam. Note that both charts show a gap where very few video lectures were watched, which is highlighted in the pictures through an oval. This gap coincides with the Dutch *Carnaval* holidays. We can observe that, in general, students that failed watched fewer video lectures. Also note that in Fig. 6.a the density of events heavily decreases after the mid-term exam (highlighted through a vertical dashed line). This could be explained by students being discouraged after a bad mid-term result. This phenomenon is also present in (b), but not equally evident. We can also observe that most students tend to constantly use video lectures. This is confirmed by the low number of students with only a few associated events.

Figure 7 shows sequence analysis models that, given any ordered sequence of activities, reflects the frequency of directly-follows relations[5] as percentage annotations

[5] The frequency of directly-follows relations is defined for any pair of activities (A, B) as the ratio between the number of times that B is directly executed after A and the total number of times that A is executed.



(a) Grade = 6 or 7 (b) Grade = 8 or 9

Fig. 7: Sequence analysis for students grouped by their course grades

and as the thickness of edges. The highest deviations from the ordered sequence order are highlighted in colored edges (i.e., black edges correspond to the natural order). This technique was tailored for the generation of reports and it is implemented using a customized RapidProM extension. When comparing (a) students that passed the course with a grade of 6 or 7 with (b) students that had a grade of 8 or 9, we can observe that both groups tend to make roughly the same deviations. Most of these deviations correspond to specific video lectures being skipped. These skipped video lectures correspond in most cases to complementary material. In general, one can observe that the thickness (i.e., frequency) of the arcs denoting the "natural" order (i.e., black arcs) is higher for (b), i.e., those with higher grades. Note that at the beginning of each week we can observe a *recovery* effect (i.e., the frequencies of the natural order tend to increase).

5.2 Summarized Report Feedback

In addition to the qualitative analysis for some courses like such as the course analyzed in Section 5.1, we have also asked lecturers for feedback through an evaluation form linked to each report.[6] The evaluation form provided 30 statements about the analysis contained in the reports (e.g., "Higher grades are associated with a higher proportion of students watching video lectures", "Video lecture views are evenly distributed throughout the course period"). Lecturers evaluated each statement on the basis of the con-

[6] The evaluation form is available at `http://www.win.tue.nl/~abolt/userfiles/downloads/Reports/form.pdf`

Table 2: Summary of the classification of statement evaluations performed by lecturers

Statement Evaluation	Core Statistics Section	Advanced Analytics Section			**Sub Total**	**Total (%)**
		Compliance	Temp. Dist.	Seq. Analysis		
Correct	261	30	67	32	**390** (89%)	**61%**
Incorrect	28	5	8	6	**47** (11%)	
Unknown	95	61	69	58	**283**	**39%**

clusions that they could draw from the report. For each of the 30 statements, lecturers could decide if they agreed or disagreed with the statement, or, alternatively, indicate that they could not evaluate the statement (i.e., "I don't know").

In total, 24 of the 89 lecturers answered the evaluation form. Out of the 720 (24 x 30) possible statement evaluations, 437 statements were answered with "agree" or "disagree". The remaining cases in which the statement could not be evaluated can be explained by three possible causes: the statement is unclear, the analysis is not understandable, or the data shows no conclusive evidence.

In the case that a statement was evaluated with "agree" or "disagree", we compared the provided evaluation with our own interpretation of the same statement for that report and classified the response as *correct* or *incorrect*. In the case that a statement was not evaluated, the respose was classified as *unknown*.

Table 2 shows a summary of the response classification for each section of the report. In total, **89%** of the statement evaluations were classified as *correct*. This indicates that lecturers were capable to correctly interpret the analysis provided in the reports. Note that the *Compliance* section had the highest rate of *unknown* classifications (63.5%). This could be related to understandability issues of the analysis presented in that section.

The evaluation form also contained a few general questions. One of such questions was: "Do you think this report satisfies its purpose, which is to provide insights about student behavior?", for which 7 lecturers answered "yes", 4 lecturers answered "no" and 13 lecturers answered "partially". All the lecturers that responded "partially" provided written feedback indicating the improvements they would like to see in the report. Some of the related comments received were: "It would be very interesting to know if students: a) did NOT attend the lectures and did NOT watch the video lectures, b) did NOT attend the lectures, but DID watch the video lectures instead, c) did attend the lectures AND watch the video lectures too. This related to their grades", "The report itself gives too few insights/hides insights", "It is nice to see how many students use the video lectures. That information is fine for me and all I need to know", and "I would appreciate a written explanation together with your diagrams, next time". Another question in the evaluation form was: "Do you plan to introduce changes in the course's video lectures based on the insights provided by this report?", for which 4 lecturers answered "yes" and 20 answered "no". The results show that the analysis is generally perceived as useful, but that more actionable information is needed, such as face-to-face lecture attendance. However, this information is currently not being recorded by the TU/e. The feedback provided by lecturers will be used to improve the report. This has only been made possible by process mining with process cubes and analytic workflows

6 Conclusion

This paper has illustrated the benefits of combining the complementary approaches of process cubes and analytic workflows in the field of process mining. In particular, the combination is beneficial when process mining techniques need to be applied on large, heterogenous event data of multidimensional nature.

To demonstrate such benefits, we applied the combined approach in a large scale case study where we provide reports for lecturers. These reports correlate the grades of students with their behavior while watching the available video lectures. Unlike existing *Learning Analytics* approaches, we focus on dynamic student behavior. Also, descriptive analytics would not achieve similar analysis results because they do not consider the process perspective, such as the ordering of watching video lectures.

Educational data has been analyzed by some disciplines in order to understand and improve the learning processes [12, 13, 14, 15, 16], even employing process cubes [17]. However, these analyses were mostly focused on individual courses. No research work has previously been conducted to allow large-scale process mining analysis where reports are automatically generated for any number of courses. Our approach has made it possible by integrating process mining with analytic workflows, which have been devised for large-scale analysis, and process cubes, which provide the capabilities needed to perform comparative analyses.

The report will be enhanced in order to incorporate the feedback obtained from lecturers through the evaluation and from now on they will be sent periodically four times per year (i.e., after each *quartile* a report is automatically generated for each course given).

The report generation will also be extended to *Massive Open Online Courses* (MOOCs) given by the TU/e. This type of courses are particularly interesting due to the fact that face-to-face lectures are not used: video lectures are the main channel used by students for accessing the course topics. For example, over **100.000** people from all over the world registered for the two executions of the MOOC *Process Mining: Data science in Action*.[7] We also plan to apply this analysis to the courses provided by the *European Data Science Academy* (EDSA).[8]

References

1. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. 1st edn. Springer-Verlag Berlin Heidelberg (2011)
2. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., van der Aalst, W.M.P.: The ProM framework: A new era in process mining tool support. In: Applications and Theory of Petri Nets. Volume 3536 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2005) 444–454
3. Jaeger, E., Altintas, I., Zhang, J., Ludäscher, B., Pennington, D., Michener, W.: A scientific workflow approach to distributed geospatial data processing using web services. In: Proceedings of the 17th International Conference on Scientific and Statistical Database Management (SSDBM'2005), Berkeley, CA, US, Lawrence Berkeley Laboratory (2005) 87–90

[7] `http://www.coursera.org/course/procmin`
[8] `http://edsa-project.eu`

4. Turner, K., Lambert, P.: Workflows for quantitative data analysis in the social sciences. International Journal on Software Tools for Technology Transfer **17**(3) (2015) 321–338
5. Bolt, A., de Leoni, M., van der Aalst, W.M.P.: Scientific workflows for process mining: building blocks, scenarios, and implementation. International Journal on Software Tools for Technology Transfer (2015) . To appear. Doi: 10.1007/s10009–015–0399–5
6. Bolt, A., van der Aalst, W.M.P.: Multidimensional process mining using process cubes. In: Enterprise, Business-Process and Information Systems Modeling. Volume 214 of Lecture Notes in Business Information Processing. Springer International Publishing (2015) 102–116
7. Cordes, C., Vogelgesang, T., Appelrath, H.J.: A generic approach for calculating and visualizing differences between process models in multidimensional process mining. In: Business Process Management Workshops. Volume 202 of Lecture Notes in Business Information Processing. Springer International Publishing (2015) 383–394
8. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.F.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **2**(2) (2012) 182–192
9. Song, M., van der Aalst, W.M.P.: Supporting process mining by showing events at a glance. In: Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS). (2007) 139–145
10. van der Aalst, W.M.P.: Process cubes: slicing, dicing, rolling up and drilling down event data for process mining. In: Proceedings of the First Asia Pacific Conference on Business Process Management. Volume 159 of Lecture Notes in Business Information Processing., Springer International Publishing (2013) 1–22
11. Mans, R.S., van der Aalst, W.M.P., Verbeek, H.M.W.: Supporting process mining workflows with RapidProM. In: Proceedings of the BPM Demo Sessions 2014 Co-located with the 12th International Conference on Business Process Management (BPM). Volume 1295 of CEUR Workshop Proceedings., CEUR-WS.org (2014) 56–60
12. Siemens, G.: Learning analytics: The emergence of a discipline. American Behavioral Scientist **57**(10) (2013) 1380–1400
13. Siemens, G., Baker, R.S.J.d.: Learning analytics and educational data mining: Towards communication and collaboration. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. LAK '12, New York, NY, USA, ACM (2012) 252–254
14. Romero, C., Ventura, S.: Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics (Part C: Applications and Reviews) **40**(6) (Nov 2010) 601–618
15. Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.: Handbook of Educational Data Mining. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press (2010)
16. Gorissen, P.J.B.: Facilitating the use of recorded lectures: analysing students' interactions to understand their navigational needs. PhD thesis, Eindhoven University of Technology (2013)
17. van der Aalst, W.M.P., Guo, S., Gorissen, P.J.B.: Comparative process mining in education: An approach based on process cubes. In: Proceedings of the 4th international Symposium on Data-driven Process Discovery and Analysis. Volume 203 of Lecture Notes in Business Information Processing., Springer Berlin Heidelberg (2015) 110–134