

# Process Mining for Recommender Strategies Support in News Media

Elena Viorica Epure, Jon Ingvaldsen, Rebecca Deneckere, Camille Salinesi

► **To cite this version:**

Elena Viorica Epure, Jon Ingvaldsen, Rebecca Deneckere, Camille Salinesi. Process Mining for Recommender Strategies Support in News Media. IEEE Tenth International Conference on Research Challenges in Information Science (IEEE RCIS 2016), Jun 2016, Grenoble, France. hal-01316566

**HAL Id: hal-01316566**

**<https://hal-paris1.archives-ouvertes.fr/hal-01316566>**

Submitted on 30 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Process Mining for Recommender Strategies Support in News Media

Elena V Epure

Universite Paris 1 Pantheon-Sorbonne  
Centre de Recherche en Informatique  
Paris, France  
Email: elena.epure@malix.univ-paris1.fr

Rebecca Deneckere

Universite Paris 1 Pantheon-Sorbonne  
Centre de Recherche en Informatique  
Paris, France  
Email: rebecca.reneckere@univ-paris1.fr

Jon Espen Ingvaldsen

Norwegian University of Science and Technology  
Department of Computer Engineering  
Trondheim, Norway  
Email: jon.espen.ingvaldsen@ntnu.no

Camille Salinesi

Universite Paris 1 Pantheon-Sorbonne  
Centre de Recherche en Informatique  
Paris, France  
Email: camille.salinesi@univ-paris1.fr

**Abstract**—The strategic transition of media organizations to personalized information delivery has urged the need for richer methods to analyze the customers. Though useful in supporting the creation of recommender strategies, the current data mining techniques create complex models requiring often an understanding of techniques in order to interpret the results. This situation together with the recommender technologies deluge and the particularities of the news industry pose challenges to the news organization in making decisions about the most suitable strategy. Therefore, we propose process mining as a high-level, end-to-end solution to provide insights into the consumers' behavior and content dynamics. Specifically, we explore if it allows news organizations to analyze independently and effectively their data in order to support them in defining recommender strategies. The solution was implemented in a case study with the third largest news provider in Norway and yielded preliminary positive results. To our knowledge, this is the first attempt to apply a process mining methodology and adapt the techniques to support media industry with the recommender strategies.

**Keywords**—process mining, inferred intentional process models, behavioral analysis, recommender strategies, process mining methodology, news media case study

## I. INTRODUCTION

In the last decades, we have witnessed a shift in the media' strategy from uniquely delivering information to the personalization of the information exposure through recommendations. As media consumers, we live in "a new age where something wonderful" - that we did not know existed - "finds us" [1]. This is accomplished resourcefully by the publishers and related organizations that have transformed the customer relationship management to a mandatory part of their business standard. For instance, looking at just two major world's leading media companies established in Europe, Le Monde Group and British Broadcasting Corporation (BBC), there is a major effort from their side in this direction. BBC integrated Sybil Recommender System<sup>1</sup> since 2012 for helping

their consumers to choose their preferred TV and radio shows, while Le Monde Group signed a long-term partnership<sup>2</sup> with Outbrain, "the world's largest content discovery platform", to improve consumers' experience in February 2015. The shift becomes prominent also through the extensive work on media-related recommender solutions in academia [2], [3], [4].

The creation of recommender solutions for *news* faces some particular challenges compared to the other types of media such as music, movies or books. The reasons are manifold: news articles loose their relevance quickly, news providers change the content often which results in a large, continuous data stream, consumers' interests are prone to be less stable [5]. In addition to this, the deluge of recommender technologies results in a strenuous process to decide on the most suitable recommender strategy for news organizations.

The recommender strategy decision requires a deep understanding of the consumers' behavior and interests [6]. Existing data mining approaches allow news organizations to gain insights into consumers from the data. Though useful, these solutions produce *very complex models* that often require thorough understanding of the techniques in order to interpret them. Yet it has been proved that to effectively support decision-making for strategy creation, the data mining techniques should be intuitive, visual, interactive and hiding the complexity [7].

In consequence, we propose *an adaption of process mining* as a high-level, end-to-end solution to reveal insights for news recommender strategies focusing in particular on the consumers' interests reflected in their behaviors. Process mining is a mature domain, originally emerging from business and information systems. Its goals are to exploit event logs captured by information systems and to answer questions such as: What happened in the past and why? What will happen in the future? When and why are there deviations

<sup>1</sup><http://www.bbc.co.uk/rd/projects/sybil-recommender-system>

<sup>2</sup><http://tinyurl.com/zpqwuvr>

from an expected process and how to control them? Process mining incorporates data mining techniques and extends them with algorithms able to mine richer models containing loops, decision points and concurrency [11]. In the current work, we extend the traditional process mining with a view focusing on the *perceived intentions* to obtain high-level models. We infer intentions from the consumers' behavior captured in database logs. However, we do not claim that these intentions are the real users' intentions without a sound experimentation in this direction. The inferred intentions in our solution play the role of a summarization mechanism, associating a semantics to the process models in order to help us and the recommender strategy' stakeholders with the analysis. These inferred intentions are linked to different news categories (e.g. "general news", "culture", "economy" etc.) and to different referrers involved in the news content dynamics (the url of the web page that contained the link the user followed to read the article; e.g. a news provider's website, social networks etc.).

The following research questions guide our research:

- Does process mining allow news organizations to exploit *independently* and *effectively* their data?
- Does process mining support news organization in defining *recommender strategies*?

To our knowledge, this is the first work to apply a solution based on process mining to data from the news industry in order to evaluate the support with recommender strategies' definition. Hence, we needed to address multiple challenges. First, the event logs we generated from the web logs needed to be enriched with a suitable semantics to enable a high-level analysis. For this purpose, we reflected on the semantics of the news data and we adopted a perspective based on inferred intention as previously explained. Second, to ensure rigor, we followed PM<sup>2</sup>, a very recent process mining methodology that yielded successful results in a case study with IBM [8]. We are the first group besides the creators that reported the use of PM<sup>2</sup> at the moment. Third, we conducted two evaluations with one organization' stakeholder: one part of the process mining methodology regarding the diagnosis and validation of the mined models, and second to evaluate the overall solution with respect to our research questions. Fourth, considering that process mining focuses also on social models, performance indicators, decision rules, and represents conformance and enhancement besides discovery [24], we analyzed how these perspectives could be exploited by the news organizations too.

The project presented in this paper was conducted as part of Smartmedia, a research program between the Norwegian University of Science and Technology (NTNU), multiple Norwegian media companies and software providers from the media sector. Considering the explorative nature of our research, we have chosen to start working with one publisher: Adresseavisen, the largest newspaper of the third largest media group in Norway (Polaris Media ASA). Nevertheless, we intend to apply the solution to the other media companies involved in the program too for an extensive evaluation. The raw data available for analysis consists of 10 millions

web events captured between August and December 2014. The results showed that process mining is able to reveal and visualize meaningful consumption patterns focusing on different perspectives corresponding to articles and sessions. Furthermore, during the evaluation, the process stakeholder acknowledged the discovery of interesting and actionable insights that could be used in the creation of recommender strategies. Some limits of the solution were uncovered too. The lack of process mining knowledge could pose issues in transforming the raw data into mineable event logs and in completely interpreting the mined processes in the beginning. However, after one hour long knowledge transfer, the process stakeholder felt confident in performing most of the process mining tasks by himself.

The paper is organized as follows. Section II introduces in more details process mining and recommender strategies and systems. Section III describes the followed methodology and the process mining approach we propose. Section IV presents the evaluation protocol and results while Section V focuses on the related work. Section VI provides a discussion about how other perspectives and types of process mining could be used for supporting the recommender strategies' creation. Section VII concludes the current work.

## II. THEORETICAL BACKGROUND

In this section, we introduce recommender systems and narrow down specifically to those focusing on the news. Further, we highlight the main goals and concepts of process mining.

### A. Recommender Systems

Recommender systems have the ability to provide relevant information to consumers based on their needs or interests. Their goal is to filter out information and present only the required and interesting items to consumers. Recommender systems can be classified into content-based, collaborative, and knowledge-based filtering [2].

Content-based filtering emerged from information retrieval and information filtering research. This method is based on item descriptions and consumers' profiles containing preferences [3]. In a content-based recommender system, features, like keywords or categories, are used to describe the items and consumers' profiles describe the type of items the consumers like. Content-based filtering tries to recommend items that are similar to those that a consumer liked in the past.

Collaborative filtering methods consist in creating consumer profiles by collecting information on consumers' behaviors and preferences, and predicting what consumers would like based on their similarity to other consumers. This method is very suitable for recommending items without understanding their content, being very suitable for movies and images [2].

A third type of recommender systems is based on explicit knowledge about items, consumer preferences, and recommendation's context. The knowledge-based recommender systems rely on a feedback loop in which the consumers requirements and preferences are gathered. When the consumers'

TABLE I  
RECOMMENDER SYSTEMS OVERVIEW [9].

Technique	Pros	Cons
Content based filtering	No knowledge from other users required, possibility to compare items	Content understanding necessary, cold start, no diversity in recommendations
Collaborative filtering	No domain knowledge needed, serendipity of results, clusters of similar users identifiable	Requires at least implicit feedback, data sparsity, cold start
Knowledge based filtering	Pre-defined recommendations, quality of recommendation, no cold-start, interactive adaptation based on a feedback loop	Knowledge engineering needed in the beginning, static knowledge, no reaction to short-term trends

preferences are hard to be described or gathered at once then the feedback loop could be employed multiple times. Furthermore, the preferences are typically not known exactly at the beginning but built within a recommendation session [3]. Table I summarizes strengths and weaknesses of these three recommender system categories.

Compared to recommender systems in media such as music, movies and books, news recommender systems face challenges because of the news industry’s particularities, including [5]

- News freshness: news articles lose their value and relevance with time, and most of the news must be presented while they are very fresh.
- Continuous news content stream: news sources publish large volumes of news articles continuously.
- Instability of consumers’ interests: the interests change much faster than in other domains (the movies preferred by a person take years to change). Consumers’ interests can change with time and contexts and consumers can also have a willingness to read news articles deviating from their common interests like breaking news.

Many practical implementations of recommender systems are hybrid systems combining elements from content-based, collaborative, and knowledge-based filtering [4]. An advantage of working with news articles is that they are a data source that machines can represent and reason about. Textual data can be broken down to a representation of significant keywords, and also linked to disambiguated semantic entities. This ability makes news data suitable for content-based filtering approaches. Many news portals also own descriptive consumer data or interaction logs. Such data allow for collaborative recommender elements with ability to identify and suggest news articles that are outside of individual common content patterns. Knowledge-based elements are also valuable for recommendation of news as they can incorporate properties such as timing (when to recommend), commercial interests and consumers’ short-term intents.

For a media company, it is a tremendous challenge to select the right recommender system. Such a strategic decision needs a deep understanding of business requirements, consumers, news contents, and their relationships. Moreover, a-priori user profiling could lead to an initial decision while posteriori knowledge could support decisions regarding adaptations and changes [1] (see Figure 1). The a-posteriori knowledge contains in fact an implicit user feedback: the captured change in reading behavior of the users resulting after the (re)deployment

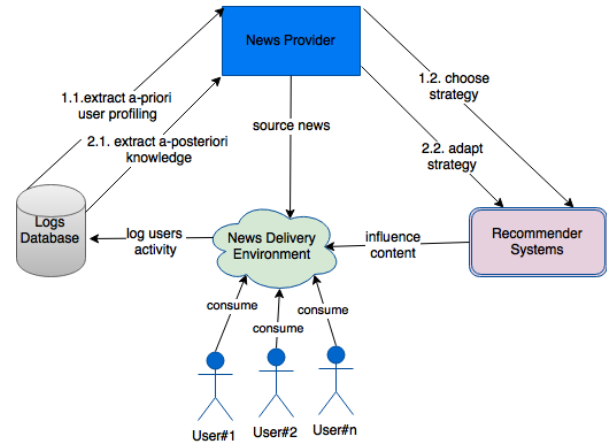


Fig. 1. News Recommendation Ecosystem.

of the recommender system [10]. The role of process mining, presented further in this section, is to support the decision-making by enabling a-priori and a-posteriori knowledge extraction (steps 1.1. and 2.1 in Figure 1). Process mining is particularly interesting in this setting as it can reveal and visualize the relationship systems, contents and users. Knowledge from such data analytics can be applied as rules and content strategies for knowledge based recommender systems and as background knowledge for designing hybrid solutions. Nonetheless, the reasoning processes over this knowledge for developing a recommender strategy is a separate research objective to our current one.

### B. Process Mining

Process mining emerged from the business and information systems community. As presented in Figure 2, process mining is a bottom-up driven approach which exploits the captured logs and existing process models: [11]

- 1) for mining process models (Discovery Process Mining);
- 2) for identifying, explaining and quantifying process differences when existing process models are compared with the discovered ones (Conformance Process Mining);
- 3) for applying changes to the existing process models based on the discovered ones (Enhancement Process Mining).

Process mining allows the analysts to get insights through mining techniques focusing on various perspectives of the same

process including time, control flow, organization, decision [11], [12]. Furthermore, simulation techniques are available for replaying the process models in order to have a deeper comprehension and capture performance indicators [13]. The extracted knowledge proved to enable discussion, decision making and hypotheses verification [11] [14].

Compared to data mining techniques such as sequence mining [15] and episode mining [16], the process mining techniques are process-driven implying the discovery of rich, high-level models with decision points, loops and concurrency. Moreover, the process models are complete, being able to capture behaviors which are also less frequent in the data [11]. Though closest to process mining, the Hidden Markov Model [17] is a time consuming, iterative procedure which produces a less accessible model to the end-user.

A specific process execution is considered a *process instance* or a *process case*. For example, if we consider a news portal, a process instance could be associated with a unique customer session. All the clicks produced within that session are named events according to the process mining terminology. These events are grouped in a *trace* which is further logged in a file. Therefore, for each process instance there is a corresponding trace in the event log file. The events in a trace are ordered by time and have multiple attributes such as the event id, the timestamp, the activity name etc. The activity names set the level of abstraction or the process granularity. At the lowest level, they could be software-generated strings such as the name of the button which was clicked. At the highest level, they could be more human-friendly such as "log in", "read front news article".

Very frequently, the event logs are generated retrospectively from databases or other types of data sources. The transformation consists in a flattening and ordering of the data where several decisions are required [11]

- Correlation decisions: What should a process instance be? What are the events that belong to a specific process instance?
- Snapshot decisions: What is the lifespan of a process instance? What should be done if a process instance is incomplete or if it is difficult to establish the end point?
- Scoping decisions: What data should be included? What attributes should be attached to each event?
- Granularity decisions: What should the process granularity be? How could it be generated automatically?

These decisions are necessary for producing event logs which represent different views of the processes. Additionally, the analyst could explore each view in different ways using the process mining tool. He could decide to observe the most frequent behavior (70% of the model for instance) or the complete one (100% of the model). He could also opt for a different meta-model for representing the process where some common possibilities are transition systems [18], Petri nets [19], Yawl [20], BPMN [21].

In our work, we adopted a process mining approach for dealing with the complexities of creating and evolving recommender strategies. Additionally, we integrated a process

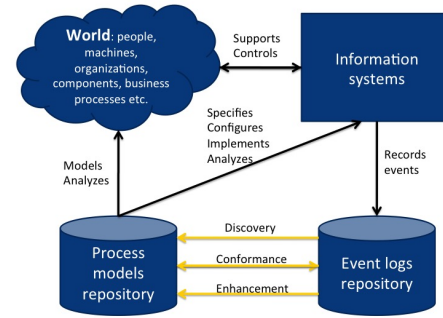


Fig. 2. Process Mining Overview. [11].

perspective based on perceived intentions in order to capture high-level, meaningful models. Previous related work proved to successfully support and guide the human agents through recommendations and predictions [22], [23].

### III. MINING PROCESSES IN NEWS MEDIA

PM<sup>2</sup>, a Process Mining Project Methodology was chosen for the systematic guidance of the solution implementation. The methodology is suited for the process mining projects that seek improvement of "process performance or compliance to rules and regulations" [8]. Though this is not our objective, we chose PM<sup>2</sup> because it is the most complete methodology compared to its counterparts - Process Diagnostics Method (PDM) and L\* life cycle model [8]. Furthermore, PM<sup>2</sup> has been already applied successfully in a case study with IBM in 2015, delivering multiple benefits. First, it could handle an extensive spectrum of process mining techniques without imposing any constraints regarding the choice of tools for executing the analysis. Second it is suited for both structured processes - those who follow a rather stable path independent from the circumstances, and unstructured processes - those who are highly flexible and unpredictable [11]. This aspect is particularly important to us, as we do not know in advance the nature of the processes that we might discover from the media web logs. Third, the methodology is organized in multiple stages for which the expected inputs, outputs and the concrete steps are specified. An overview of the methodology execution in our case is provided in Figure 3. Each of the stages is discussed further.

#### A. Planning

The purpose of the solution was established: *to discover high-level process models regarding the news consumers' behavior and the content dynamics from the media organization's data*. The consumers' interests were acknowledged to constantly change over time and contexts [5]. For example, a person could become suddenly more interested in the "culture" news under the influence of a new group of friends. Hence, we were interested in the content dynamics too as we claim it is less sensitive to changes in the long term. If positive, this

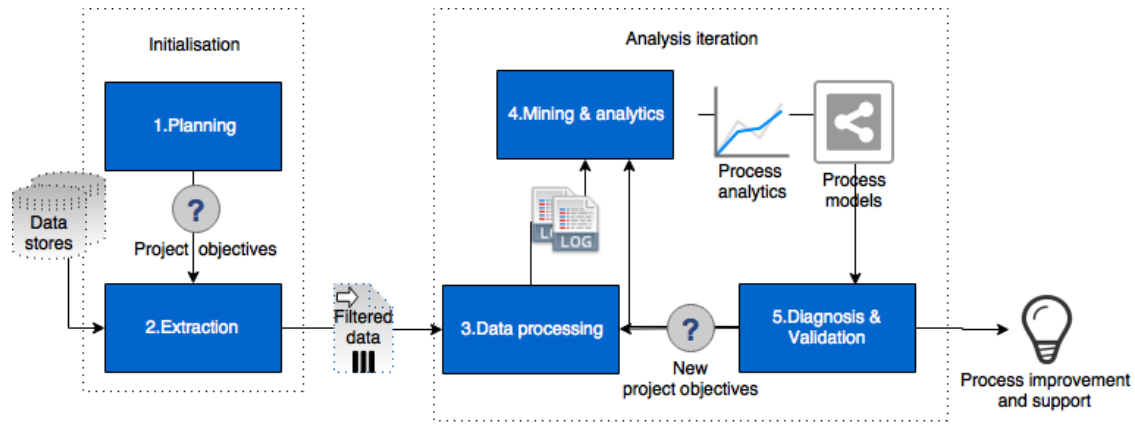


Fig. 3. Process Mining Project Methodology [8].

could represent a steadier and more reliable input to create recommender strategies taking in consideration the balance of long versus short consumers' interests. For instance, knowing the life-cycle of "culture" articles and how this type of content propagate over various referrers in time could help the news provider to make decision about when and where to deliver this content.

The accuracy of the processes is directly linked to the data quality. Process mining starts from the assumptions that the data is representative and must capture significant behavior. In reality, there are two problems that could appear: 1) noise when the data captures a rare or infrequent behavior; 2) incompleteness when the data is not sufficient to discover a structure [24]. These potential issues were handled by working with data spanning several months and by injecting a suitable semantics in the event logs as described in Section III-C.

### B. Data Extraction

The second stage implied the extraction and integration of data from different systems. For starters, the scope was determined: data spanned the period August - December 2014 and represented information about clicks and articles. For each click the following attributes were extracted: the session id, the article id, the referrer and the elapsed time since the article's publication. Furthermore, the extracted attributes for the articles were: the id, the category, the address of the webpage, the publication's timestamp and some other metrics such as the average lifetime, the total number of views, the number of views on the publisher website or originating from Facebook or external sources. In total 10 million clicks regarding 3427 articles were extracted. Knowledge was transferred tacitly during this phase through data comprehension. This guided us further in understanding how we could generate the event logs in order to reveal the target processes.

### C. Data processing

In this stage, event logs corresponding to our target processes were generated. We built the event logs around a semantics that could capture the customers' behavior and the

content dynamics. In order to do so, we set up two experiments that are described further.

1) *Article-centered Experiment*: The first experiment was built starting from the following scenario. An article is initially published on the news provider's website. During its lifespan the article is accessed through different sources marked by the referrer attribute. We decided to associate each article with a process instance and each referrer with an activity in order to obtain process models regarding the content's dynamics. The media organization could perform various process mining activities regarding these models:

- *Discovery*: How do articles propagate over sources in time? How do articles from specific categories propagate over various sources in time? How do articles with a different level of popularity propagate over various sources in time?
- *Comparison*: Do articles coming from different categories propagate in a similar way over various sources in time? Do articles having a different level of popularity propagate in a similar way over various sources in time?
- *Prediction*: How could the news' provider expect that an article coming from a specific category that exceeded a certain threshold in popularity will propagate further?

When we analyzed the data we discovered there were 928 referrers. A process with so many different activities would have been unfeasible for analysis or decision-making support. However, we noticed these referrers embedded knowledge that could help us with the interpretation of the users' intention through their semantics. For example, when a person clicked on an article shared on social media, we considered that person followed the community recommendation. If another person accessed the same article after a search query, we considered that he was searching for news. Enriching events with such semantics not only allowed us to perform aggregation but also to create simplified, high-level process models.

Consequently, we opted for the as-is aggregation where each click was associated to a more general class: the perceived intention corresponding to the semantics of its referrer. The further challenges were the identification of the intentions



TABLE II  
THE MOST POPULAR REFERRERS AND THEIR ASSOCIATED INTENTIONS.

Referrer	Url	Clicks	Intention	Description
Adressa	www.adressa.no	7158398	Read_News_Provider	article on the news provider's website
.	Unknown	1359473	-	the referrer is not known thus nor the intention
Facebook	m.facebook.com	460287	Follow_Comm_Recm	the Facebook community's recommendation
Sol	www.sol.no	213163	Follow_Aggeg_Recm	the Sol news aggregator's recommendation
Facebook	www.facebook.com	98764	Follow_Comm_Recm	the Facebook community's recommendation
Disqus	redirect.disqus.com	89001	Follow_Comm_Recm	the Disqus community's recommendation
Sol	tablet.sol.no	79097	Follow_Aggeg_Recm	the Sol news aggregator's recommendation
Facebook	l.facebook.com	64034	Follow_Comm_Recm	the Facebook community's recommendation
Sol	m.sol.no	61272	Follow_Aggeg_Recm	the Sol news aggregator's recommendation
Google	www.google.no	31628	Search_News	article found by the Google search engine
Disqus	disqus.com	20649	Follow_Comm_Recm	the Disqus community's recommendation
Facebook	lm.facebook.com	17494	Follow_Comm_Recm	the Facebook community's recommendation
Sol	app.sol.no	13277	Follow_Aggeg_Recm	the Sol news aggregator's recommendation
Twitter	t.co	6415	Follow_Comm_Recm	the Twitter community's recommendation
Aftenposten	www.aftenposten.no	4489	Follow_Aggeg_Recm	the Aftenposten news aggregator's recommendation
Startsiden	www.startsiden.no	3639	Follow_Aggeg_Recm	the Startsiden news aggregator's recommendation
Google	www.google.com	2629	Search_News	article found by the Google search engine
Google	news.google.com	1872	Follow_Aggeg_Recm	the Google news aggregator's recommendation
Vg	www.vg.no	1764	Follow_Aggeg_Recm	the Vg news aggregator's recommendation
Mosjoen	www.mosjoen.com	1705	Follow_Aggeg_Recm	the Mosjoen news aggregator's recommendation
Vg	direkte.vg.no	1573	Follow_Aggeg_Recm	the Vg news aggregator's recommendation
Mosjoen	mosjoen.com	1218	Follow_Aggeg_Recm	the Mosjoen news aggregator's recommendation

TABLE III  
EXAMPLES OF RULES FOR IDENTIFYING THE INTENTION.

if ('search' in Referrer_Url) or ('google' in Referrer_Url and 'news.google' not in Referrer_Url) then Search_News
if ('facebook' in Referrer_Url) or ('twitter' in Referrer_Url) then Follow_Comm_Recm

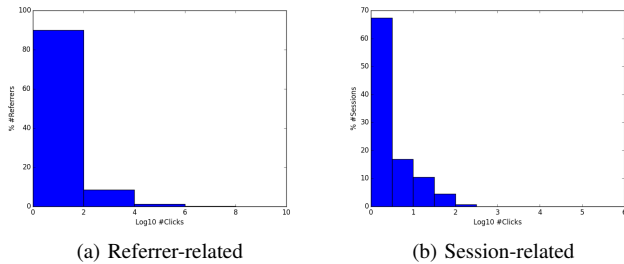


Fig. 4. Histograms of Clicks in Log10 Scale.

covering the data set and the tagging. The former was done manually while the later automatically. Looking at the distribution of the number of clicks per referrer in Figure 4, we observed that approximately 90% of the referrers had less than 100 clicks associated. If 900 from 928 referrers generated in average 50 clicks this meant that less than 0.5% clicks from the whole data set was associated with the majority of referrers (45000 clicks of 10 million). We decided thus to safely skip the intention identification for these referrers as their small proportion of generated clicks appeared to not impact the results. The method for identifying the intentions consisted in the following steps:

- 1) Discover the most popular referrers having at least 1000 clicks associated.
- 2) Define perceived intention for each popular referrer (Table II).

- 3) Create a set of rules for tagging the clicks based on Table II and other observed common patterns in the data (some examples in Table III).
- 4) Apply the rules automatically to the data set.

Intentions could be more complex, when they contain sub-intentions, or simpler. In this situation we opted for simple, perceived intentions through a naive interpretation of the referrers' semantics. We avoided to make any claim about more complex intentions. However, even in this case false positives might exist. For example someone could click on a news link from his Facebook page by mistake. These cases could be discovered often by analyzing extra data as the time spent on reading the article.

Finally, we also decided to filter the logs in order to obtain different perspectives on the data and reduce complexity [8]. The methods employed were "slice and dice" and "compliance based filtering". Slice and dice was performed first for creating different views for low-popularity, medium-popularity and high-popularity articles; and for different categories of articles. Compliance based filtering was performed for every different view by removing the events for which the referrer was unknown or had less than 1000 clicks associated.

2) *Session-centered Experiment*: The second experiment was centered on the sessions. Each session represented a process instance and the article's category was associated with an activity. The media organization could perform various process mining activities regarding the customers' behavior with a focus on the evolution of their reading interests.

- *Discovery*: How do the interests of customers evolve over time? What types of content drive and form longer engagement cycles?
- *Comparison*: Are there any differences in the processes for long versus short sessions?
- *Prediction*: For a specific session, what is the type of article that would most likely be read next?

Similar to the previous experiment, we filtered the logs applying the two methods: slice and dice, and compliance based filtering. The former method was chosen to create different views for long versus short sessions. There were 539120 sessions having the distribution presented in Figure 4. The decision to classify a session in long or short was made based on the number of clicks per session; the limit was chosen to be 30 in order to have a sufficient number of instances in the long sessions group. Though, other thresholds should be experimented in the future.

#### D. Mining and analysis

Stage 4 consisted in applying process mining and analytics techniques to the event logs received from the previous stage in order to gain insights into the customers' behavior and content dynamics. The event logs were created as XES files [11] by using a set of scripts written in Python<sup>3</sup> on the original data. We performed the analysis using Disco<sup>4</sup>, a commercial suite supported by the leading academic group in process mining at Eindhoven Institute of Technology. Though not as rich in techniques as ProM [25], Disco has proved to be user friendly and very fast on big event logs while discovering high-level knowledge. Moreover, it has been successfully used in numerous industry case studies and academic projects<sup>5</sup>. The exact process mining algorithms used in Disco are not public but the owners claim that a technique based on Fuzzy Miner [26] is implemented. Currently, we have not tested any other process mining technique apart from this one but we plan it in our future work.

We will expose only a part of the findings to exemplify how the process mining tool could be used and how to interpret the mined processes. A node in the graph represents an activity while the directed edge represents the flow between the activities. The numbers associated to both nodes and edges give information about the frequencies. This also influences the color of the nodes, respectively the width of the edges. Two special nodes exist: the start and the stop.

A process could be observed at different level of details. For instance, Figure 5 and Figure 6 represent the same process corresponding to the high-popularity articles in the article-centered experiment. The former could be useful for providing an accurate overview of the process metrics while the later exposes the most common behavior. In addition, as the second version (Figure 6) represents a simpler view of the process, it is much easier to compare it to other process models as for

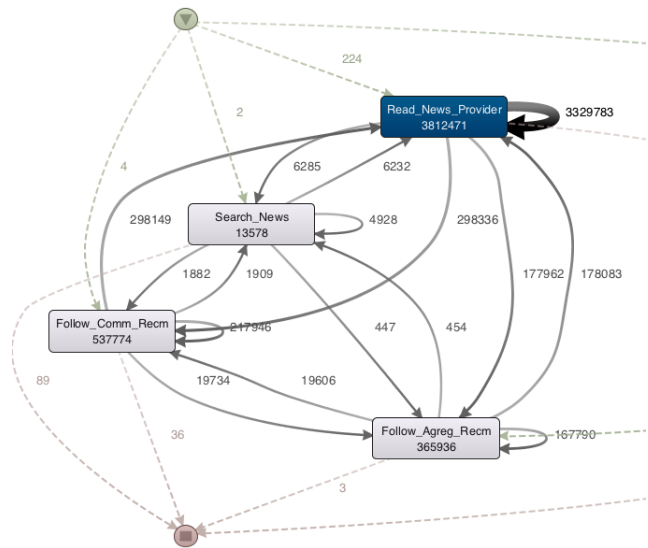


Fig. 5. Article-Centered Process High Popularity Articles - 100% Details.

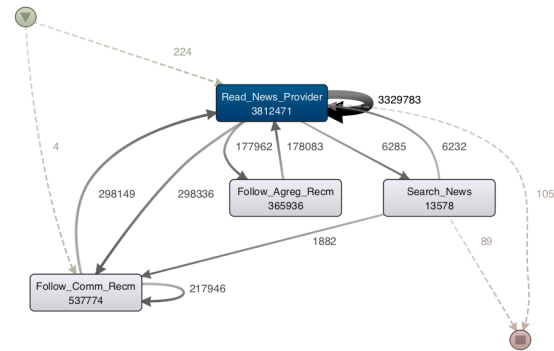


Fig. 6. Article-Centered Process High Popularity Articles - 50% Details.

example the process corresponding to the low-popularity article (Figure 7). The processes for both low and high-popularity articles are very similar. It could be noticed that towards the end of its life, an article is accessed by consumers mainly from two sources: the news provider website or through search engines. Another common aspect is the fact that the second popular referrer for the articles after the news provider website is the social media. This could be observed based on the width of the process edges. There are though several differences in the two processes: once a low-popularity article is read by consumers through search engines, it continues to be accessed in this way (the self loop of the Search\_News activity in Figure 7); same observation stands for high-popularity articles and the social media (the self loop of the Follow\_Comm\_Recm activity in Figure 6). We performed the process comparison manually given the manageable level of details but automatic techniques for model comparison are available in ProM [25].

Regarding the session-centered experiments, a fragment of the discovered process models corresponding to the short

<sup>3</sup><https://www.python.org>

<sup>4</sup><https://fluxicon.com/disco/>

<sup>5</sup><http://fluxicon.com/blog/>



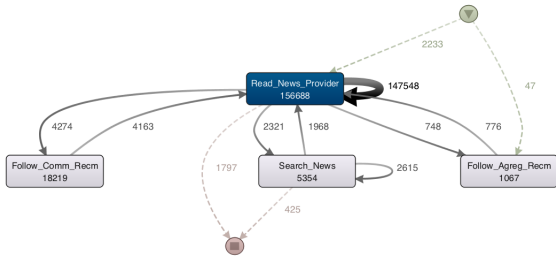


Fig. 7. Article-Centered Process Low Popularity Articles - 50% Details.

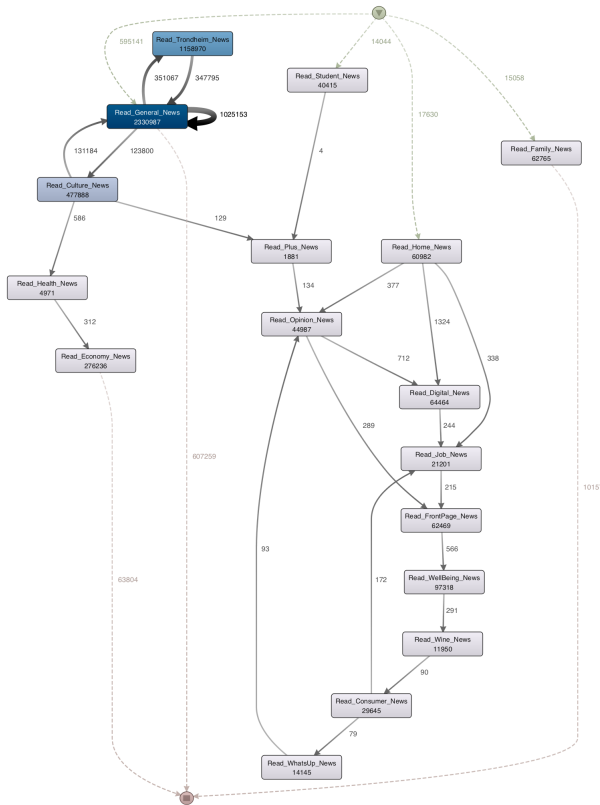


Fig. 8. Session-Centered Process Short Sessions - 10% Details.

session is presented in Figure 8. It is very interesting to notice that the most read categories are General news, Trondheim news and Culture news (the color of the boxes is more intense if the activity appears more often). Moreover, these consumers' interests seem to cluster: the consumers reading General news most probably read Trondheim news and Culture news too. On the contrary, Family news are more likely read in isolation from the other categories.

### E. Diagnosis and Validation

According to the PM<sup>2</sup> methodology, a process mining project should end with the diagnosis and validation of the findings [8]. This phase targets the process stakeholders.

Process mining research states that given sufficient data, the techniques could accurately reflect the logs [24], [26]. Consequently, as we worked with an already proven algorithm (Disco's Fuzzy miner extension [26]), we considered the mined processes correct. The objectives for the diagnose phase were thus to understand the mined process models, interpret the results correctly, identify interesting or peculiar aspects, and propose research questions for further iterations. The objective for validation was to compare the findings to the expectations of the process stakeholder.

This phase was conducted as a semi-structured interview with Jørgen Frøland, the Project Manager in User Experience and Big Data of the case media group. He has been responsible for the application of recommender systems for more than one year. For this reason, we considered him as the process stakeholder, the role with the more extensive knowledge about the news consumers and recommender strategies within the organization.

The process stakeholder acknowledged an easier interpretation and understanding of the models regarding the session-centered experiment while for the second he requested a more detailed explanation of the process semantics. He observed that different behavioral patterns appeared repeatedly across multiple models as expected. On the contrary, some peculiar results lead to valuable insights such as adding more promote links (Plus news) in the content of the Culture news than in other types. Additionally, the consumption patterns of Economy news were found also very interesting as it seemed the consumers were less likely to jump to other succeeding categories. Another unexpected observation was the fact that Opinion news seemed to be a bridge or access point to other news categories.

The main future objectives mentioned during the interview were the thorough validation of these models and to explore the extent to which the current approach could measure the impact of a recommender strategy. The process stakeholder said: "We as an organization need to establish a trust relationship to this kind of models [...] If I starts recommending Consumer article in Wine articles, will this increase the engagement on the site?". Other types of experiments including the temporal aspect and animations for re-playing process instances over a time period were of particular interest too.

Finally, the process stakeholder found the mined process models very meaningful and aligned with their previous analysis. What was considered especially valuable was the flow between various perceived intentions and interests. The process stakeholder concluded "it met my expectations very well [...] I could sit hours playing with the complexity sliders, take notes and explore the data."

## IV. PRELIMINARY EVALUATION

This work is an exploratory study where the use of process mining to support the definition of recommender strategies was investigated. The nature of the research was influenced by the fact that no process mining solution has been applied before to web logs originating from the news industry. Aligned with

the research questions, the solution evaluation aimed at the verification of the following claims:

- 1) An end-to-end tool mining high-level processes could support the process stakeholder of the news organization to perform the tasks independently because the tool is easy to use (visual, intuitive, hiding the complexity of the mining techniques).
- 2) The mined processes and the process mining tool provide meaningful insights to the process stakeholder into the consumers' behavior and content dynamics.
- 3) The discovered insights could support the process stakeholder in defining recommender strategies through predictive analysis and decision-making support.

The intuition behind these claims emerged from the authors' experience with process mining and was backed-up by the existing research in the area.

The solution evaluation started with an experiment that allowed the subject to perform process mining using Disco. The event logs were prepared in advance by the authors. An introduction of the process mining concepts and the prepared event logs was given to the subject in the beginning. Following the experiment, a questionnaire containing mainly open-ended questions was used during a semi-structured interview. Additional questions and probing beyond the original ones were welcomed. The researcher took notes during the experiment and recorded the interview. The overall evaluation lasted three hours. For data analysis, we employed inductive analysis, which derives general patterns from the transcripts' detailed information [27]. Finally, the interpretations were validated with the subject.

The subject highlighted several benefits of the process mining tool. First, the visualization and the overall user interface seemed to be very well perceived especially considering the organization's background - "we are not hard core programmers, we use technology services and SaaS solutions.". The solution was intuitive as there were "not so many feature options" and the model changed immediately when interacting with the tool. Moreover, the subject strongly agreed that the complexity of the mining techniques was hidden. However, one of challenges stated by the subject was to get acquainted with the process mining concepts and to understand the relationship between event logs and processes. Additionally, the interpretation of the process models was also challenging in the beginning because of the learning curve regarding the process model representation (understanding what the boxes, edges, numbers attached to the edges etc. represent). This was also caused by the fact the models did not go always from upper left corner to the bottom right one and the layout was re-shaped every time when the complexity thresholds were re-adjusted.

Regarding the process models, the subject acknowledged that they gave him interesting insights into the consumers' behavior and content dynamics. Also, he stated the organization could act upon the discovered knowledge. The subject found "interesting to use process mining in a continuous digital strategy" as this would allow to measure and monitor the effects of the taken actions. The decision-making support of

the solution was confirmed and two possible use cases were highlighted by the subject: to assist "discussions on how to handle personalization vs. the storage of personal sensitive information." and to drive sales strategies. With respect to prediction analysis, the subject found the current discovered knowledge insufficient and required more process models spanning various seasons and days.

In conclusion, process mining seems to allow independent use after the understanding of the process mining concepts, event logs and the tool. The subject found the mined knowledge and the tool very promising for creating recommender strategies.

Validity threats were handled during the research to ensure the rigor and quality of the results. According to Yin there are four types of validity that must be considered in a case study: construct, internal, external and reliability [28]. In order to ensure this, multiple measures were taken. Preliminary interviews were conducted with the company to better understand the situation. Several control actions were performed: the subject was informed about the evaluation procedures in advance, a case study protocol was followed, the experiment and the interviews were recorded in real time. Additionally, data triangulation was performed through the inclusion of multiple evidence sources: primary documents belonging to the news organization, secondary documents from the existing body of knowledge in recommender systems and process mining, direct observations during the experiment and interviews' transcripts. Nevertheless, there are some threats to validity that should be elaborated. First, as the preliminary evaluation took place with only one subject, the results might have been sometimes biased by the subject's attitude towards the experiment. For example, the excitement to discover a new tool might yield, maybe unconsciously, a rather positive feedback during the interview. Second, there is a researchers' bias in generating the log which comes from the perceived intentions. False positives could exist even though we take extra measures such as using the reading duration. Moreover, the set of perceived intentions we chose should be validated in the future with real users too.

Regarding the external validity that concerns the generalization of the results, though the solution was applied to a single case study, some of its characteristics are generic. News logs generally contain timestamps and information regarding the categories and referrers. Hence, with respect to applying process mining, the current case settings are representative for the news industry. However, differences could appear in the usefulness' perception of process mining results for recommender strategy. The stakeholders could have various technical or application domain expertise, objectives or expectations. Therefore, generalization could not be claimed yet.

## V. RELATED WORK

The current work could be considered an application of web usage mining [29] with the purpose of creating recommender strategies. Web usage mining is concerned with the discovery and prediction of the users' behavior interacting with the

world wide web from the log records. Compared to techniques employed in web mining - sequential patterns, clustering, classification, hidden Markov chains [17], we propose a new approach based on process mining that discovers high-level process models.

There are few published works within the scope of web mining who employed process mining: in e-commerce [30] and in e-learning [31]. To our knowledge, there is no previous attempt in the media domain. The trigger of these works regards the strategy improvement of the organizations through competitive advantage in the education system [30] and business efficiency in the e-commerce systems [31]. Similar to our work, the authors claim benefits of process mining in this direction through the discovery of high-level behavioral models. The benefits come in the form of support for decision making, recommendations and prediction. Contrary to our approach, these works do not follow any process mining methodology and exploit less events than we do: 8 thousands events [31] respectively 4 million events [30]. Moreover, regarding the models, [31] analyses especially the social perspective while the [30] adopt control-flow perspective with mixed semantics mainly referring to the website navigation. We make a clear separation of the activities' meanings by mapping them on the consumers' perceived intentions and consumers interests.

Most of the existing research in web mining apply traditional data mining techniques [32], [33], [34], [35], [36]. The objectives of these works are to improve web site usages [32], [33] and to create recommender systems [34], [35]. Ontologies are also used to model the human behavior when navigating in information networks [36].

Regarding the applications to media industry, web mining is employed for gaining insights into the consumers' reading preferences for recommendation and engagement [37], [38], [39], [40], [41]. However, the behavior is modeled or analyzed differently in all these works: [37] discover a referrer graph, [38] investigate the effect of news polarity on the users gaze, affect and attention, [39] mine transition probabilities of moving from one news category to another, and [40] identify different reading behaviors using a method based on matrix factorization. In our solution, we exploit both the news referrers and news categories while the identification of distinctive behaviors is implicit through event log filtering and variations in process mining analysis. Finally, while some of the media-related works exploit real data [38], [40] like us, others use data collected from existing data sets [39] or lab experiments [38], [41].

## VI. PERSPECTIVES

In this work, the mined models represent an activity-centered, control flow perspective of the processes. However, existing process mining techniques could create complementary models focusing on other aspects of the processes. A process case could have different attributes. These attributes can be exploited by the techniques to provide a "helicopter view" of the event log known as dotted chart [42], [11]. For our first experiment centered on articles, some case attributes

that should be further explored for generating the dotted chart are: the article category and the article popularity. Similarly, for our second experiment centered on sessions, a possible case attribute could be the consumer's location.

Then, organizational process mining aims at creating social networks with the resources involved in the process [24]. Usually, the resources are people but other mappings are possible too. This type of mining allows the discovery of profiles: resources that are involved in similar activities in the same order. An interesting future direction in our work would be to consider the referrers being resources for the session-centered processes. In this way, we could observe if some referrers are similar in terms of promoted news categories which are more likely to be read from their web pages and the reading order.

Further, performance process information could be mined too. For example, the average time between two activities could be embedded into the model. Another possibility would be to project time on resources. For example, the news organization could analyze the reading time between different news categories. Forth, decision rules could be learned from event logs and further used for recommendation and prediction [43].

As we mentioned in the Section II, process mining could be performed for conformance checking too [11] [13]. This supports the automatic comparison of the process models. In the future, it would be very interesting to compare process models which are mined for different periods such as seasons or days of the week. An additional comparison we would want to perform is with the results obtained for other news providers. Conformance checking could also have a significant contribution in measuring the impact of recommender strategies. Currently, the evaluation methods rely on technical measures such as click through rates, average read time of articles, customer return rates [44]. It could be useful to complement this approach with one involving the comparison of process models before and after the adoption or change of recommender strategies.

Previous work regarding recommendation techniques for resource allocation in process mining could be also applied [45]. For example, for the session-centered experiment, if referrers are resources, these techniques could make predictions if an article for which the category is known, will be clicked within that session if exposed by a specific referrer. The prediction algorithm takes into account the profiles of the resources.

Finally, process mining solutions for dealing with continuous streams of data should be also considered [46]. This would allow the process discovery and update in real time while the clicks are recorded. All these process mining perspectives and types are tool-supported [25].

## VII. CONCLUSION

A formal definition of data scientist still faces difficulties to emerge from the plethora of descriptions, situations and sources. However, this is of less importance compared to the drivers of such a job. In the pursuit of solving problems or

improving services, any role, from human resources to marketing or management, could qualify as a data scientist when empowered with the right tools and experience. Nonetheless, while in real life business stakeholders sometimes possess the right tools, a situation frequently encountered is the one where they struggle with the complexity [7].

In this work, we explored for the first time process mining as a means to allow the news organizations to independently and meaningfully exploit their data for consumers' insights and recommender strategies' creation. As mentioned in Section II, there are several challenges regarding the media sector: 1) news freshness, 2) continuous news content streaming and 3) the instability of consumers' interests. Recommender strategies have implemented solutions to handle the first and second issues by considering the news timestamps and popularity, and by evolving the recommendation models as new user data was logged [47]. Process mining could be especially useful as support in the offline investigation of consumers' interests. The changes produced in the consumers' reading behavior could be analyzed with process mining tools independently by the organization, given the graphical, high-level representation.

The solution was implemented in a case study with a Norwegian news provider. The preliminary evaluation results were very encouraging. Process mining proved effective in understanding news readers behavior. Furthermore, it was perceived by professionals as bringing more thoughtful knowledge than activity monitoring on the relationship between the consumers' behavior and recommended contents. However, few limits also need further tackling in a more extensive evaluation, planned with other news providers.

Recommender systems exist to provide to the consumers the most appropriate and relevant content. This could only be possible by having a rich understanding of how people act and their underlying goals. In future work, we will tackle this more thoroughly by applying new types and perspectives of process mining.

## REFERENCES

- [1] V. Moscato, A. Picariello, and A. M. Rinaldi, "Towards a user based recommendation strategy for digital ecosystems," *Knowledge-Based Systems*, vol. 37, pp. 165 – 175, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950705112002092>
- [2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 6, pp. 734–749, 2005.
- [3] L. Chen, M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, and G. Semeraro, "Human decision making and recommender systems," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 3, no. 3, p. 17, 2013.
- [4] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, no. 4, pp. 331–370, 2002.
- [5] M. Harandi and J. A. Gulla, "Survey of user profiling in news recommender systems," *3rd International Workshop on News Recommendation and Analytics (INRA 2015), Vienna, Austria*, 2015.
- [6] P. Seitlinger, D. Kowald, C. Trattner, and T. Ley, "Recommending tags with a model of human categorization," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2013, pp. 2381–2386.
- [7] T. C.-K. Huang, C.-C. Liu, and D.-C. Chang, "An empirical investigation of factors influencing the adoption of data mining tools," *International Journal of Information Management*, vol. 32, no. 3, pp. 257 – 270, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0268401211001319>
- [8] M. van Eck, X. Lu, S. Leemans, and W. van der Aalst, "Pm2: A process mining project methodology," in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, J. Zdravkovic, M. Kirikova, and P. Johannesson, Eds. Springer International Publishing, 2015, vol. 9097, pp. 297–313. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-19069-3\\_19](http://dx.doi.org/10.1007/978-3-319-19069-3_19)
- [9] D. Jannach and G. Friedrich, "Tutorial: Recommender systems," in *Proceedings of the International Joint Conference on Artificial Intelligence, Barcelona*, 2011.
- [10] D. Parra, A. Karatzoglou, I. Yavuz, and X. Amatriain, "Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping," in *In Proceedings of the ICARIS-2011*, 2011.
- [11] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer Publishing Company, Incorporated, 2011.
- [12] A. Tiwari, C. Turner, and B. Majeed, "A review of business process mining: state-of-the-art and future trends," *Business Process Management Journal*, vol. 14, no. 1, pp. 5–22, 2008. [Online]. Available: <http://dx.doi.org/10.1108/14637150810849373>
- [13] A. Adriansyah, J. Munoz-Gama, J. Carmona, B. van Dongen, and W. van der Aalst, "Measuring precision of modeled behavior," *Information Systems and e-Business Management*, vol. 13, no. 1, pp. 37–67, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s10257-014-0234-7>
- [14] W. van der Aalst, "Extracting event data from databases to unleash process mining," in *BPM - Driving Innovation in a Digital World*, ser. Management for Professionals, J. vom Brocke and T. Schmiedel, Eds. Springer International Publishing, 2015, pp. 105–128. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-14430-6\\_8](http://dx.doi.org/10.1007/978-3-319-14430-6_8)
- [15] S. Parthasarathy, M. J. Zaki, M. Ogihara, and S. Dwarkadas, "Incremental and interactive sequence mining," in *Proceedings of the Eighth International Conference on Information and Knowledge Management*, ser. CIKM '99. New York, NY, USA: ACM, 1999, pp. 251–258. [Online]. Available: <http://doi.acm.org/10.1145/319950.320010>
- [16] W. Zhou, H. Liu, and H. Cheng, "Mining closed episodes from event sequences efficiently," in *Advances in Knowledge Discovery and Data Mining*, ser. Lecture Notes in Computer Science, M. Zaki, J. Yu, B. Ravindran, and V. Pudi, Eds. Springer Berlin Heidelberg, 2010, vol. 6118, pp. 310–318. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-13657-3\\_34](http://dx.doi.org/10.1007/978-3-642-13657-3_34)
- [17] L. Rabiner and B. Juang, "An introduction to hidden markov models," *ASSP Magazine, IEEE*, vol. 3, no. 1, pp. 4–16, Jan 1986.
- [18] T. Henzinger, Z. Manna, and A. Pnueli, "Timed transition systems," in *Real-Time: Theory in Practice*, ser. Lecture Notes in Computer Science, J. de Bakker, C. Huizing, W. de Roever, and G. Rozenberg, Eds. Springer Berlin Heidelberg, 1992, vol. 600, pp. 226–251. [Online]. Available: <http://dx.doi.org/10.1007/BFb0031995>
- [19] M. Hack, "Petri net language," Cambridge, MA, USA, Tech. Rep., 1976.
- [20] W. van der Aalst and A. ter Hofstede, "Yawl: yet another workflow language," *Information Systems*, vol. 30, no. 4, pp. 245 – 275, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306437904000304>
- [21] J. Recker, "Opportunities and constraints: the current struggle with bpmn," *Business Process Management Journal*, vol. 16, no. 1, pp. 181–201, 2010. [Online]. Available: <http://dx.doi.org/10.1108/14637151011018001>
- [22] E. Epure, C. Hug, R. Deneckère, and S. Brinkkemper, "What shall i do next?" in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, M. Jarke, J. Mylopoulos, C. Quix, C. Rolland, Y. Manolopoulos, H. Mouratidis, and J. Horkoff, Eds. Springer International Publishing, 2014, vol. 8484, pp. 473–487. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-07881-6\\_32](http://dx.doi.org/10.1007/978-3-319-07881-6_32)
- [23] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi, "Process mining versus intention mining," in *Enterprise, Business-Process and Information Systems Modeling*, ser. Lecture Notes in Business Information Processing, S. Nurcan, H. Proper, P. Soffer, J. Krogstie, R. Schmidt, T. Halpin, and I. Bider, Eds. Springer Berlin Heidelberg, 2013, vol. 147, pp. 466–480. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-38484-4\\_33](http://dx.doi.org/10.1007/978-3-642-38484-4_33)

- [24] W. van der Aalst and M. Song, "Mining social networks: Uncovering interaction patterns in business processes," in *Business Process Management*, ser. Lecture Notes in Computer Science, J. Desel, B. Pernici, and M. Weske, Eds. Springer Berlin Heidelberg, 2004, vol. 3080, pp. 244–260. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-25970-1\\_16](http://dx.doi.org/10.1007/978-3-540-25970-1_16)
- [25] B. van Dongen, A. de Medeiros, H. Verbeek, A. Weijters, and W. van der Aalst, "The prom framework: A new era in process mining tool support," in *Applications and Theory of Petri Nets 2005*, ser. Lecture Notes in Computer Science, G. Ciardo and P. Darondeau, Eds. Springer Berlin Heidelberg, 2005, vol. 3536, pp. 444–454. [Online]. Available: [http://dx.doi.org/10.1007/11494744\\_25](http://dx.doi.org/10.1007/11494744_25)
- [26] C. W. Günther and W. M. P. Aalst, *Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ch. Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, pp. 328–343. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-75183-0\\_24](http://dx.doi.org/10.1007/978-3-540-75183-0_24)
- [27] B. M., R. J., and M. L., *Real World Evaluation: Working under budget, time, data, and political constraints*. Sage Publications, 2006.
- [28] R. K. Yin, *Case study research*. Sage Publications, 2003.
- [29] K. Sharma, G. Shrivastava, and V. Kumar, "Web mining: Today and tomorrow," in *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, vol. 1, April 2011, pp. 399–403.
- [30] N. Poggi, V. Muthusamy, D. Carrera, and R. Khalaf, "Business process mining from e-commerce web logs," in *Proceedings of the 11th International Conference on Business Process Management*, ser. BPM'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 65–80. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-40176-3\\_7](http://dx.doi.org/10.1007/978-3-642-40176-3_7)
- [31] H. Awatef, B. Gueni, M. Fhima, and S. David, "Process mining in the education domain," *International Journal on Advances in Intelligent Systems*, vol. 8, no. 1-2, 2015.
- [32] P. Thomas, "Using interaction data to explain difficulty navigating online," *ACM Trans. Web*, vol. 8, no. 4, pp. 24:1–24:41, Nov. 2014. [Online]. Available: <http://doi.acm.org/10.1145/2656343>
- [33] M. Spiliopoulou, C. Pohle, and L. Faulstich, "Improving the effectiveness of a web site with web usage mining," in *Web Usage Analysis and User Profiling*, ser. Lecture Notes in Computer Science, B. Masand and M. Spiliopoulou, Eds. Springer Berlin Heidelberg, 2000, vol. 1836, pp. 142–162. [Online]. Available: [http://dx.doi.org/10.1007/3-540-44934-5\\_9](http://dx.doi.org/10.1007/3-540-44934-5_9)
- [34] R. Bhushan and R. Nath, "Automatic recommendation of web pages for online users using web usage mining," in *Computing Sciences (ICCS), 2012 International Conference on*, Sept 2012, pp. 371–374.
- [35] M. Hosseinzadeh Aghdam, N. Hariri, B. Mobasher, and R. Burke, "Adapting recommendations to contextual changes using hierarchical hidden markov models," in *Proceedings of the 9th ACM Conference on Recommender Systems*, ser. RecSys '15. New York, NY, USA: ACM, 2015, pp. 241–244. [Online]. Available: <http://doi.acm.org/10.1145/2792838.2799684>
- [36] D. Lamprecht, M. Strohmaier, D. Helic, C. Nyulas, T. Tudorache, N. F. Noy, and M. A. Musen, "Using ontologies to model human navigation behavior in information networks: A study based on wikipedia," *Semantic Web*, vol. 6, no. 4, pp. 403–422, 2015.
- [37] M. Trevisiol, "Exploiting implicit user activity for media recommendation," *SIGIR Forum*, vol. 49, no. 1, pp. 70–70, Jun. 2015. [Online]. Available: <http://doi.acm.org/10.1145/2795403.2795421>
- [38] I. Arapakis, M. Lalmas, B. B. Cambazoglu, M.-C. Marcos, and J. M. Jose, "User engagement in online News: Under the scope of sentiment, interest, affect, and gaze," *Journal of the Association for Information Science & Technology*, vol. 65, no. 10, pp. 1988–2005, October 2014. [Online]. Available: <https://ideas.repec.org/a/bla/jinfst/v65y2014i10p1988-2005.html>
- [39] C. Esiyok, B. Kille, B.-J. Jain, F. Hopfgartner, and S. Albayrak, "Users' reading habits in online news portals," in *Proceedings of the 5th Information Interaction in Context Symposium*, ser. IiX '14. New York, NY, USA: ACM, 2014, pp. 263–266. [Online]. Available: <http://doi.acm.org/10.1145/2637002.2637038>
- [40] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi, "Improving user topic interest profiles by behavior factorization," in *Proceedings of the 24th International Conference on World Wide Web*, ser. WWW '15. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2015, pp. 1406–1416. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2736277.2741656>
- [41] A. M. Jorge, J. P. Leal, S. S. Anand, and H. Dias, "A study of machine learning methods for detecting user interest during web sessions," in *Proceedings of the 18th International Database Engineering & Applications Symposium*, ser. IDEAS '14. New York, NY, USA: ACM, 2014, pp. 149–157. [Online]. Available: <http://doi.acm.org/10.1145/2628194.2628239>
- [42] J. Claes, I. Vanderfeesten, J. Pinggera, H. Reijers, B. Weber, and G. Poels, "Visualizing the process of process modeling with ppmcharts," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, M. La Rosa and P. Soffer, Eds. Springer Berlin Heidelberg, 2013, vol. 132, pp. 744–755. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-36285-9\\_75](http://dx.doi.org/10.1007/978-3-642-36285-9_75)
- [43] A. Rozinat and W. van der Aalst, "Decision mining in prom," in *Business Process Management*, ser. Lecture Notes in Computer Science, S. Dustdar, J. Fiadeiro, and A. Sheth, Eds. Springer Berlin Heidelberg, 2006, vol. 4102, pp. 420–425. [Online]. Available: [http://dx.doi.org/10.1007/11841760\\_33](http://dx.doi.org/10.1007/11841760_33)
- [44] P. Cremonesi, F. Garzotto, and R. Turrin, "User-centric vs. system-centric evaluation of recommender systems," in *Human-Computer Interaction – INTERACT 2013*, ser. Lecture Notes in Computer Science, P. Kotze, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, Eds. Springer Berlin Heidelberg, 2013, vol. 8119, pp. 334–351. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-40477-1\\_21](http://dx.doi.org/10.1007/978-3-642-40477-1_21)
- [45] M. Arias, E. Rojas, J. Munoz-Gama, and M. Sepulveda, "A framework for recommending resource allocation based on process mining," in *3th International Workshop on Decision Mining and Modeling for Business Processes (DeMiMoP&AZ15)*, August-September 2015.
- [46] A. Burattin, A. Sperduti, and W. van der Aalst, "Control-flow discovery from event streams," in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, July 2014, pp. 2420–2427.
- [47] A. Lommatzsch and S. Albayrak, "Real-time recommendations for user-item streams," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, ser. SAC '15. New York, NY, USA: ACM, 2015, pp. 1039–1046. [Online]. Available: <http://doi.acm.org/10.1145/2695664.2695678>