# CJM-ex: Goal-oriented Exploration of Customer Journey Maps using Event Logs and Data Analytics

Gaël Bernard<sup>1</sup> and Periklis Andritsos<sup>2</sup>

<sup>1</sup> University of Lausanne, Faculty of Business and Economics (HEC), Switzerland gael.bernard@unil.ch
<sup>2</sup> University of Toronto, Faculty of Information, Canada periklis.andritsos@utoronto.ca

Abstract. Customer Journey Mapping (CJM), is an emerging area of research tackling issues related to customer behavior and user trajectories when consuming a service. The increasing complexity of the service industry makes this type of tools popular amongst practitioners. However, to date, it is not clear how a CJM can be used to depict hundreds or thousands of customer journeys. Inspired by process discovery techniques – borrowed from Process Mining – we present *CJM-explorer* (CJM-ex). CJM-ex is a web interface that uses hierarchical clustering and statistical indexes to allow interactive navigation, with or without a-priori information, through numerous journeys stored in standard event log formats. The exploration of the underlying journeys can be done in the whole set of data available or driven by user goals in order to examine events and patterns in specific areas of interest.

**Keywords:** customer journey mapping, process mining, customer journey analytics, hierarchical clustering, sequence mining

## 1 Motivations

In order to deliver great service, companies need to have an understanding of the quality of customer experience at an end-to-end level, [5,7]. The ever growing amount of services offered to users for consumption has made the ability to understand their behavior very important, [8]. Similarly important is the knowledge extracted by the increasing number of ways organizations interact with their customers; e.g., a customer might visit a physical store, purchase a product online, and provide feedback on social media. As a response, new customer-centric approaches have surfaced. The *customer journey map* (CJM) is an example of such new techniques. CJMs allow for better understanding of a customer's end-to-end experience when using a service by mapping any interactions with the company (called touchpoints) on a map that ultimately contributes to better understanding and serving customer needs.

In our previous work, [4], we discussed the benefit of bringing customer journey mapping and process mining together by proposing a CJM model that can be used with process mining techniques. More specifically, CJM-ex aims at providing a solution to explore numerous customer journeys at the same time. Similar to discovery techniques used in process mining, our algorithm takes event logs as input, without using any a-priori information, [1]. However, instead of outputting a business process model (BPM), we display the journeys onto a CJM. For a complete definition of process mining and event logs, the reader is referred to the well known Process Mining Manifesto, [2]. Visualizing event logs on CJMs, instead of BPMs, exhibits two interesting features. First, CJMs focus more on personal customer activities (e.g., by incorporating customer emotions), rather than the "internally-focus problem-solving approach" of BPMs, [3]. Second, contrary to BPMs, CJMs can incorporate customer journeys that are deemed exceptional behaviors, rather than removing them to increase model readability.

Despite these interesting features, representing many customer journeys onto a CJM in an intelligible manner remains a challenge. Current research typically limits the number of journeys to be compared to less than ten, making the overall process relatively straightforward. However, we argue that companies in the service industries tend to deal with hundreds or thousands of journeys. To overcome this challenge and identify different areas of interest, a hierarchical clustering algorithm is employed to segment the original data. The hierarchical nature allows for a top-down navigation of automatically generated groups of similar journeys. Once the clusters are formed, CJM-ex is able to leverage the contextual information that comes along a typical customer journey such as the customers' characteristics, or the emotions, [4]. It does so in two different ways. First, we employ statistical indexes in order to explain why the different clusters were generated. Second, we let the users define their own exploration goals, making CJM-ex the first goal-oriented tool that allows analysts to set a-priori goals to guide their journey exploration.

We implemented CJM-ex, which aims to: 1) show how numerous event logs can be displayed onto CJMs; and 2) let users navigate into these journeys. The next section introduces our tool, while the third section highlights the key parts of its technical implementation. We conclude by providing an outlook.

## 2 CJM-ex

The main objective of CJM-ex is to let users upload and explore their own dataset using a customer journey map layout. To limit the number of journeys displayed on the same CJM and allow their intuitive exploration, we took a hierarchical clustering approach, as illustrated in Fig. 1. Each letter represents an activity and



Fig. 1. Hierarchical clustering of journeys

the tree is built bottom up by merging the activities that are most similar at each iteration. By default, our hierarchical algorithm uses a proximity measure that takes into account the order of activities and is a variant of the Jaccard similarity based on shingles, [6]. In our clustering tree (or dendrogram), the journeys seen in the event logs are at the leaf level and as they get merged they form "representative" journeys. A representative journey is a single pattern of activities whose purpose is to summarize the patterns contained in a cluster. Because the representative journeys at the top of the tree summarize many – potentially distant – journeys, it will tend to show only few activities shared by many. In contrast, the representative journeys closer to the leafs will show more details. Borrowing a cartographic metaphor from [1], the first layers show general patterns and hide less important activities – like a world map would omit small cities. However, our application allows a drill-down into 'countries' of interest (i.e., pattern of activities), which would redirect to new CJMs where the previously hidden 'cities' (i.e., omitted activities) will be shown.

CJM-ex is accessible at http://customer-journey.unil.ch, where we also provide a screencast explaining its usage. The screenshot visible in Fig. 2 points out three views that are available to navigate the clusters. Each of these views fullfills specific objectives. First, the CJM view **0** shows journeys that are in the same cluster. This representation allows to easily compare the pattern of activities. Second, the tree **2** displays the hierarchical structure of the journey clusters – useful in providing a holistic



Fig. 2. Interface pointing to three views: O CJM,Ø tree and, Ø textual boxes.

view of the clusters and where we currently are. Third, a box per cluster ③ provides a convenient means to display statistical indexes that we named "salient characteristics". The salient characteristics is the top 5 results of a chi-square test applied on all the contextual information. For instance, if at a global level (the entire dataset) the number of women is equal to the number of men, it might be surprising to find a cluster with large majority of women. Therefore, this information might come up as one of the top 5 salient characteristics.

Moreover, the user might be interested in specific characteristics occurring during the journey. For this reason, we allow user-defined goals. For instance, one might be interested in journeys that started by the activity "attending class" experienced by young people. The top part ① of Fig. 3 displays the settings, while the bottom part ② shows that some branches of the tree are interesting with regards to the goal (red "hot" area at the top). Hence, our application allows navigation without using any a-priori information, but also setting navigation goals, and guidance by the resulting colors.



Fig. 3. ① goal settings, ② impact on the tree.

Finally, when moving from one view to the others, the three views are updated synchronously, allowing a smooth exploration amongst journeys.



## 3 Implementation

CJM-ex is build around four main elements: 1) a web interface; 2) the XES-parser; 3) Hcluster; and 4) a data warehouse. We will describe the main parameters and choices we made for each of them.

Web interface. The web interface leverages bootstrap<sup>3</sup>, jquery<sup>4</sup> and  $d_{3js}^{5}$  to provide a user-friendly interface to upload and navigate journeys. Both the CJM view and the tree view are implemented in d<sub>3js</sub>. The CJM view is our own implementation, while the tree is an adaptation of existing code<sup>6</sup>.

**XES Parser.** CJM-ex works with event logs. More specifically, we leverage the XES (eXtensible Event Stream) standard born within the process mining taskforce. The XES Parser is a Java implementation that encapsulates the OpenXES library<sup>7</sup> to parse XES file. Use of this open source software ensures that our application is strictly compatible with the XES standard.

**Hcluster.** Hcluster is a python implementation containing the three steps illustrated in Fig. 5. The first one is the hierarchical clustering implemented using Scipy<sup>8</sup>. Two parameters should be provided as inputs: the distance measure between event sequences and the methods for calculating the distance between clusters. They can both be chosen by the user when uploading a dataset (see Fig. 4). While further research is needed to understand why these different field.



Fig. 4. Preview of the parameters when uploading a dataset

ferences exist, it seems that using shingles, [6], as a distance metric provides a more intuitive way to navigate through journeys. Once the hierarchical cluster is formed, the next step consists of cutting the clustering to form layers. A *layer* is a set of predetermined number of journeys that will be grouped together and will ultimately appear on the same CJM. To achieve this, we developed an algorithm that recursively cuts the dendrogram returned by Scipy. A small number of journeys will lead to a simple CJM that is easy to visualize, but in more complex tree structures (i.e., trees with larger height). Finally, the last step consists of finding the representative journey using a frequent sequences mining algorithm<sup>9</sup>.

**Data Warehouse.** Each dataset is saved in its own database schema designed as a star schema. Due to lack of space, we do not include its full schema, which

- <sup>8</sup> https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html
- <sup>9</sup> https://github.com/bartdag/pymining/blob/master/pymining/seqmining.py

<sup>&</sup>lt;sup>3</sup> http://getbootstrap.com/

<sup>&</sup>lt;sup>4</sup> https://jquery.com/

<sup>&</sup>lt;sup>5</sup> https://d3js.org/

<sup>&</sup>lt;sup>6</sup> http://bl.ocks.org/robschmuecker/7880033

<sup>&</sup>lt;sup>7</sup> http://www.XES-standard.org/openXES/start

stores all the information required to use the application (e.g., clusters, journeys, events) as well as some precomputations. For instance, we count the number of occurences for each characteristic at each cluster, so the goals and the salient characteristics can be retrieved quickly.

Altogether, the parameters visible in Fig. 4 allow users to explore the exact same dataset from different perspectives.

## 4 Discussion and outlook

With CJM-ex, we have demonstrated that many journeys can be displayed onto CJM in an intelligible and efficient manner, offering an alternative to a BPM representation. As Gartner highlighted, CJMs should be used to complement, but not to replace, BPMs, [9]. By leveraging a standard born within the process mining taskforce (i.e., XES) and by mimicking a typical process mining activity (i.e., discovery), we bring these techniques closer together. However, further research is required to fully understand how they can complement each other. This is a call to the process mining and business process management communities to consider CJMs as an integral part of the process management toolkit.

#### References

- 1. van der Aalst, W.: Process Mining: Data Science in Action. Springer (2016)
- van der Aalst, W., Adriansyah, A., De Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., et al.: Process mining manifesto. In: International Conference on Business Process Management. pp. 169–194. Springer (2011)
- van den Bergh, J., Işik, O., Viaene, S., Helsen, E.: Re-positioning business process management: Exploring key capabilities for successful business transformation (2016)
- 4. Bernard, G., Andritsos, P.: A process mining based model for customer journey mapping. In: Proceedings of the Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017) (2017)
- 5. Bernard, G., Boillat, T., Legner, C., Andritsos, P.: When sales meet process mining: A scientific approach to sales process and performance management. In: International Conference on Information Systems (2016)
- Broder, A.Z., Glassman, S.C., Manasse, M.S., Zweig, G.: Syntactic clustering of the web. Computer Networks and ISDN Systems 29(8-13), 1157–1166 (1997)
- Haugstveit, I.M., Halvorsrud, R., Karahasanovic, A.: Supporting redesign of c2c services through customer journey mapping. In: Service Design Geographies. Proceedings of the ServDes. 2016 Conference. pp. 215–227. No. 125, Linköping University Electronic Press (2016)
- 8. Lemon, K.N., Verhoef, P.C.: Understanding customer experience throughout the customer journey. Journal of Marketing pp. jm–15 (2016)
- Olding, E., Cantara, M., Robertson, B., Dunie, R., Huang, O., Searle, S.: Predicts 2016: Business transformation and process management bridge the strategytoexecution gap. Tech. rep., Gartner (November 2015), https://www.gartner.com/ doc/3173020/predicts--business-transformation-process