



# Discovering Customer Journeys from Evidence: a Genetic Approach Inspired by Process Mining

# CAISE'19

Gaël Bernard, University of Lausanne, Faculty of Business and Economics (HEC), Switzerland  
Periklis Andritsos, University of Toronto, Faculty of Information (iSchool), Canada

## ABSTRACT

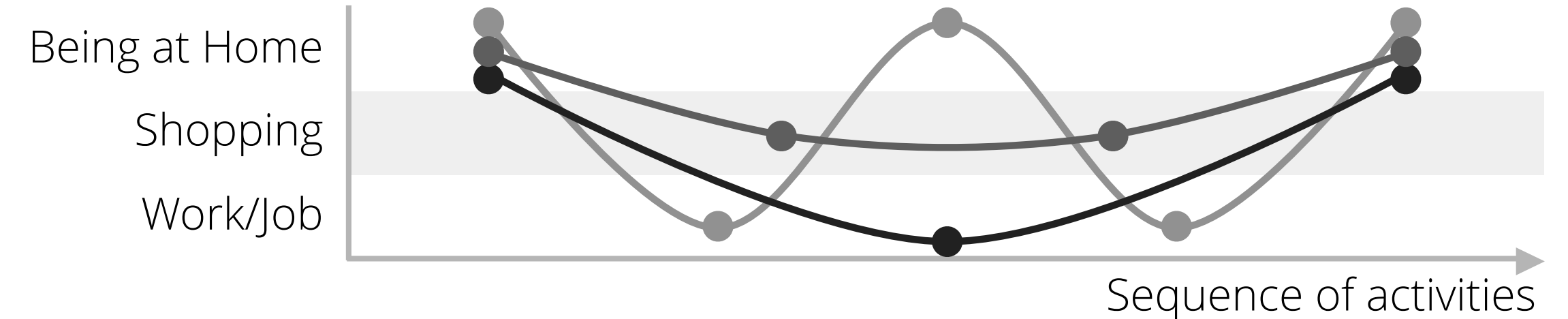
By depicting the main customer trajectories, a *Customer Journey Map* (CJM) helps service providers put themselves in their customers' shoes. However, existing CJMs are built by hand by internal stakeholders, resulting in subjective CJMs. Inspired by Process Mining techniques, we propose an innovative approach to build *evidence-based CJM*.

Input

XES

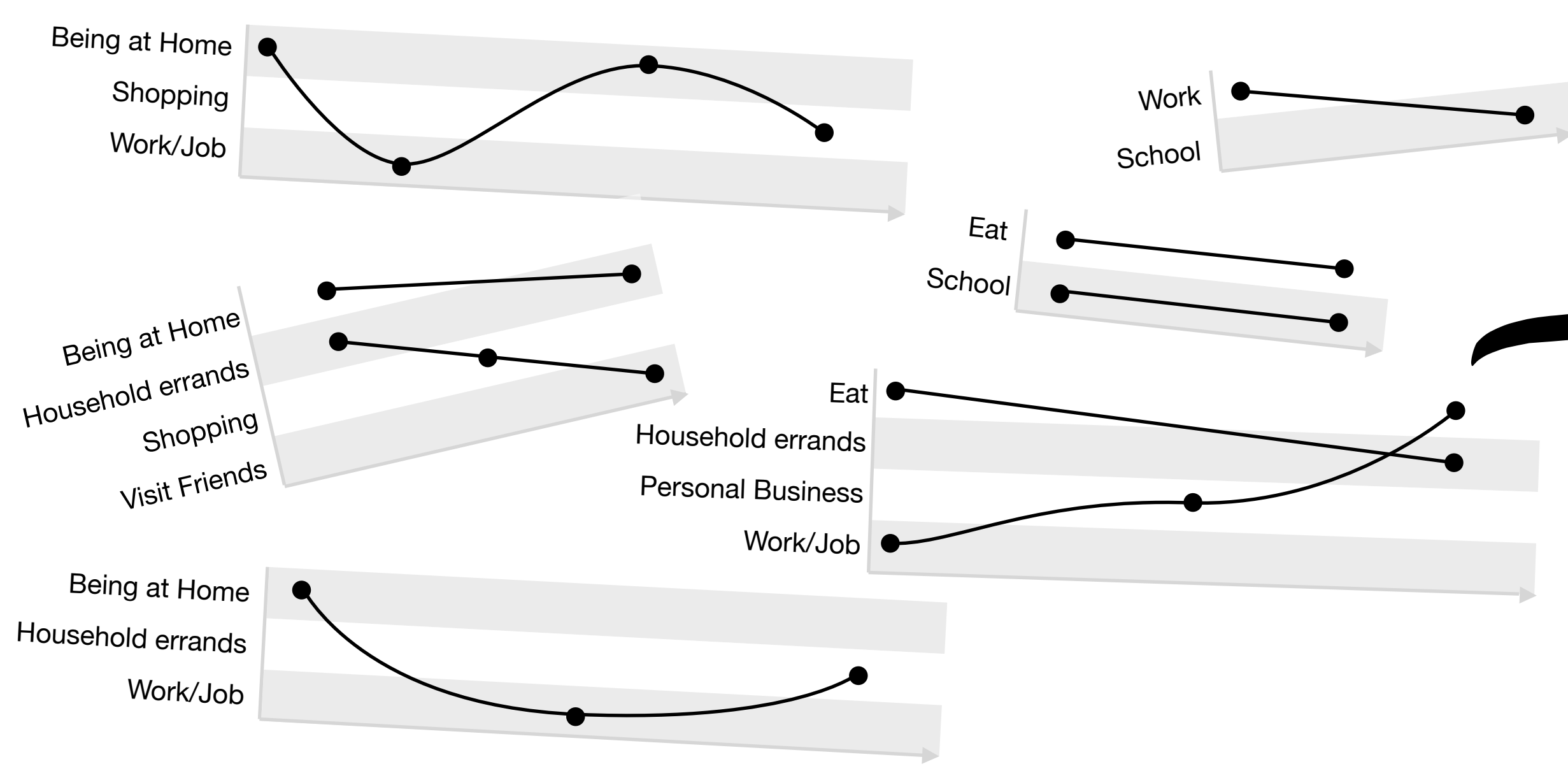
We store customer journeys using the eXtensible Event Stream (XES) standard.

Output



PROPOSED GENETIC PROCESS

## 1. Generate a Population of (random) CJMs



## 2. Evaluate the CJMs

Step 1: Assign each trace to its closest journey.  
Step 2: Average the following qualities.

$$\text{Fitness}(J_a, J_{\mathcal{R}}) = 1 - \frac{\sum_{i=1}^{|J_a|} \min_{j=1}^{|J_{\mathcal{R}}|} (\text{Levenshtein}(\sigma_{A_i}; \sigma_{R_j}))}{\sum_{i=1}^{|J_a|} \text{Length}(\sigma_{A_i})}$$

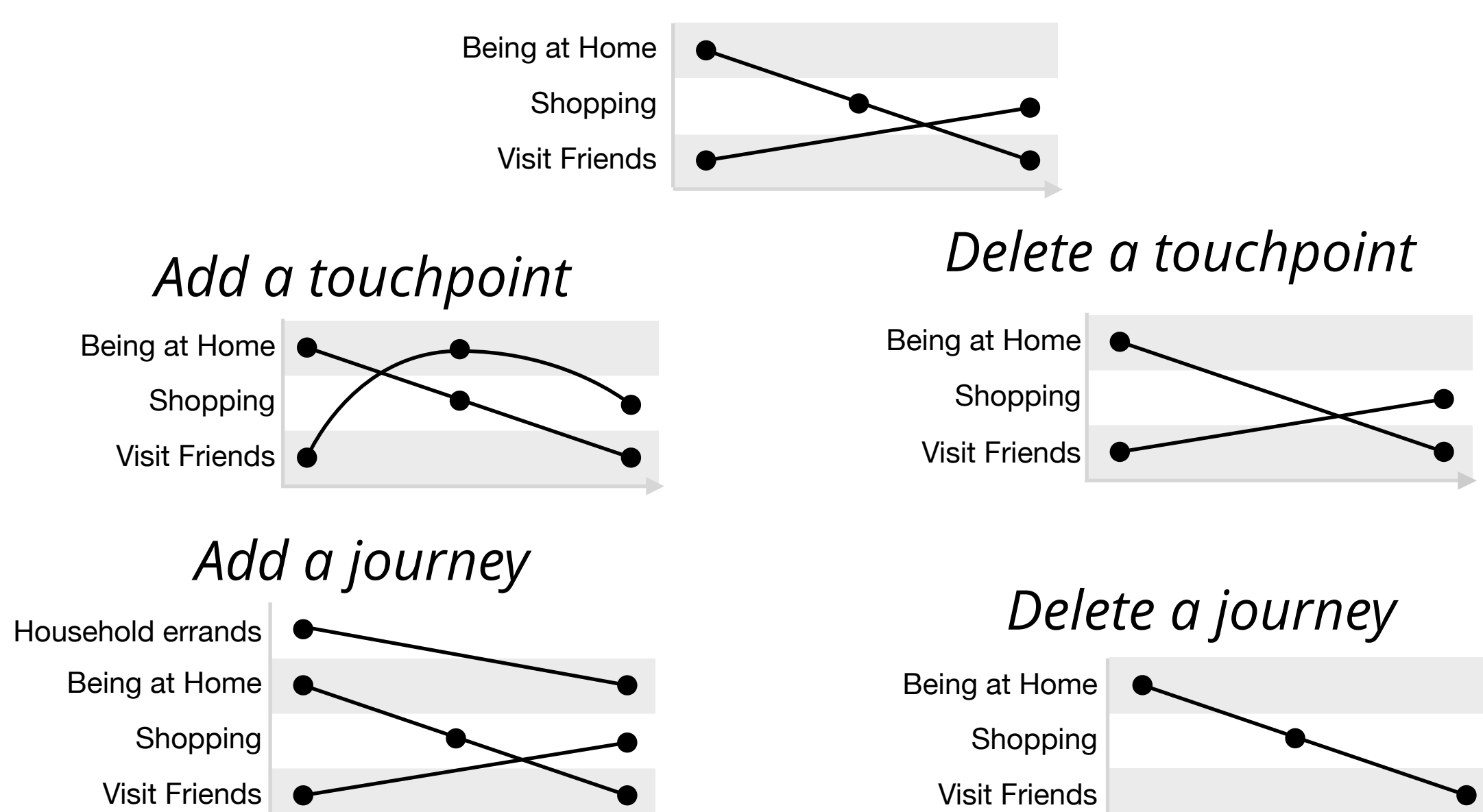
Measures the distance between each trace and its closest representative using the Levenshtein distance

$$\text{NumberOfRepresentatives}(k_{\mathcal{R}}, k_h, x_0) = \frac{1}{1 + (\frac{|k_{\mathcal{R}} - k_h|}{x_0})^2}$$

Penalizes solutions that are far from the "optimal" number of clusters.

## 4. Transform the CJMs

We apply one of the following transformations to create new CJMs: (Original)



## 3. Termination criteria met?

- 1) Maximum number of generations reached  
Typically 500 generations.
- 2) The quality does not improve anymore  
Based on fitness and number of representatives. We fix the limit to 50 generations in our experiments.

no

yes, return best CJM

## EVALUATION

- We build datasets to represent customer journeys  
They are found under: <http://customer-journey.unil.ch/datasets/>
- We compare with...  
...an approach combining the Calinski-Harabasz index with *Kmedoids*  
...techniques from social sciences; i.e., *Traminer* [1]
- Jaccard-index w.r.t. original datasets. 1=perfect match.

Genetic:	Traminer:	Kmedoids:
0.656	0.579	0.462

### Levenshtein Mean-distance

between the representative sequence x and the sequence of actual journeys that have been assigned to x. 0=perfect match

Genetic:	Traminer:	Kmedoids:
1.37	1.72	1.75

## OUTLOOK

- CJM can be built from evidence  
Similar to how Business Process Models can be discovered by Process Mining.
- The work can be extended  
e.g., by considering customers' demographic information.
- Interdisciplinary Research Required  
Research at the crossroads between data science, marketing, and social sciences is key to a full understanding of customer experiences

### References:

[1] Gabadinho, A. & Ritschard, G. (2013), Searching for typical life trajectories applied to child birth histories, In R. Lévy & E. Widmer (eds.), *Gendered life courses*. pp. 287-312. Vienna: LIT

Link to the paper:

