ETH zürich

A Deep Generative Model for Clickstream Analysis

Yilmazcan Ozyurt, Tobias Hatt, Ce Zhang, Stefan Feuerriegel
ETH Zurich

AMLD EPFL 2022, 28th March 2022

ETH zürich

Overview

Problem

- In 2020, over 97% of global online shopping users **exited** the e-commerce websites **with no purchase.**
- E-commerce websites use **machine learning** to **detect when users are about to exit,** which helps them to trigger potential interventions:
  - Special coupons
  - Dynamic price promotions

- **Research question**: How can we detect the users who will exit with no purchase?

Previous Works

Table 1: Overview of key literature for predicting user exits without purchase from clickstreams

Model	Long-term dependence	Latent states
LSTM [16, 32]	✓	
BiLSTM [39]	✓	
Mixture LSTM [31, 36]	✓	
HMM [5, 24]		✓
M3PP [11]		✓

- Variant of recurrent neural networks (RNN) to capture **long-term dependence** of the sequence of page clicks.
- Variant of Markov models to capture the **latent states** (i.e. latent shopping phases) of the users.

Overview

Problem

- In 2020, over 97% of global online shopping users **exited** the e-commerce websites **with no purchase.**
- E-commerce websites use **machine learning** to **detect when users are about to exit,** which helps them to trigger potential interventions:
  - Special coupons
  - Dynamic price promotions
- **Research question**: How can we detect the users who will exit with no purchase?

Our approach

- We formalize a **tailored attentive deep Markov model** called **(*Clickstream*DMM)** for detecting user exits without purchase.
- Our model is the first clickstream model that jointly learns both
  - **long-term** dynamics of the page clicks (via attention)
  - different **shopping phases** in clickstream setting (via a latent variable model).
- Our *Clickstream*DMM consistently **outperformed** state-of-the-art algorithms in terms of
  - AUROC of 0.817 (**+11.5 %** improvement)
  - Based on **26,279 online sessions** provided by **Digitec Galaxus** (leading e-commerce platform in Switzerland).

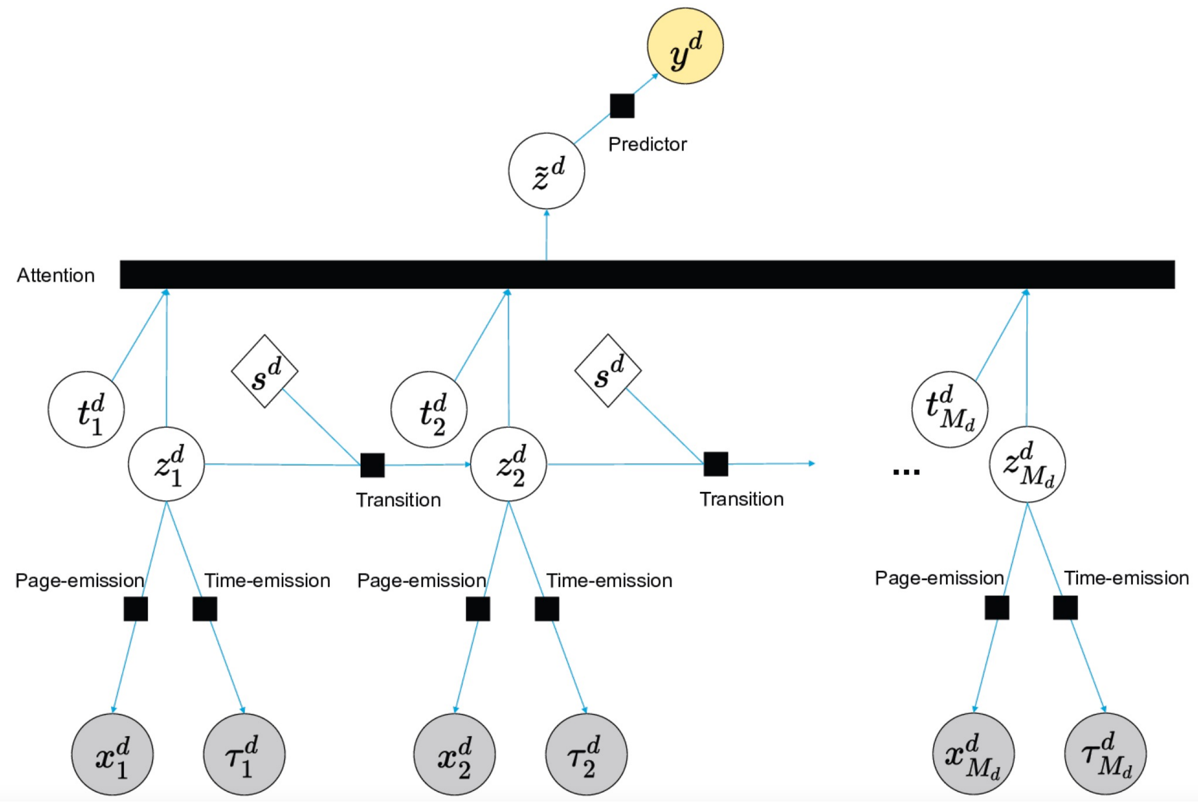
Clickstream Data and Model

Clickstream Data

- The dataset comprises a set of online sessions; each sessions is represented by **static** and **time-series** features.
  - **Static features** encode user-level information (e. g., gender, age, account type), denoted by s.
  - **Time-series** features encode the set of **page clicks** $\{x_m\}_{m=1}^{M}$ and their associated **timestamps** $\{t_m\}_{m=1}^{M}$.
    - We further derive **time spent on a page** (TSP) $\{\tau_m\}_{m=1}^{M}$.
- For each online sessions, we predict whether the user will end the session with vs. without purchase $y \in \{0,1\}$.
  - $y = 0$ denotes no purchase,
  - $y = 1$ denotes purchase before the exit.

Model

Figure 1: Model Architecture of *Clickstream*DMM. Black squares denote neural networks.



Clickstream Data and Model

Components

- **Transition network** specifies transition probability among consecutive latent variables.

$$p(z_m \mid z_{m-1}, s)$$

- **Page-emission network** outputs the probability of a page click given the latent variable.

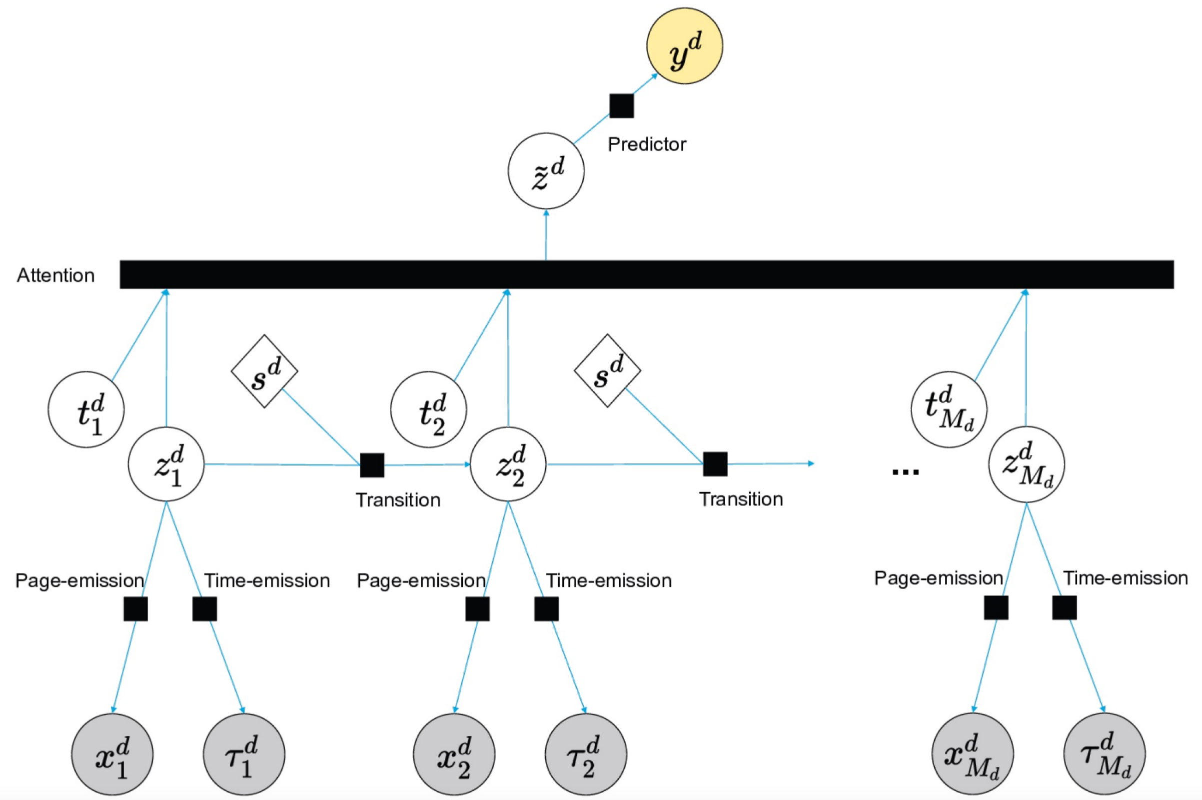
$$p(x_m \mid z_m)$$

- **Time-emission network** outputs the probability of time spent on the page (TSP) given the latent variable.

$$p(\tau_m \mid z_m)$$

Model

Figure 1: Model Architecture of *Clickstream*DMM. Black squares denote neural networks.



Clickstream Data and Model

Components

- **Attention network** allows the model to assign a different importance to each latent variable when making inferences of purchase probability.

$$z'_m = \tanh(W_{z'}\, z_m + b_{z'}),$$

$$\gamma_m = \frac{\exp(K(v,\, z'_m))/(t_M + 1 - t_m)^\beta}{\sum_{m'=1}^{M} \exp(K(v,\, z'_{m'}))/(t_M + 1 - t_{m'})^\beta},$$

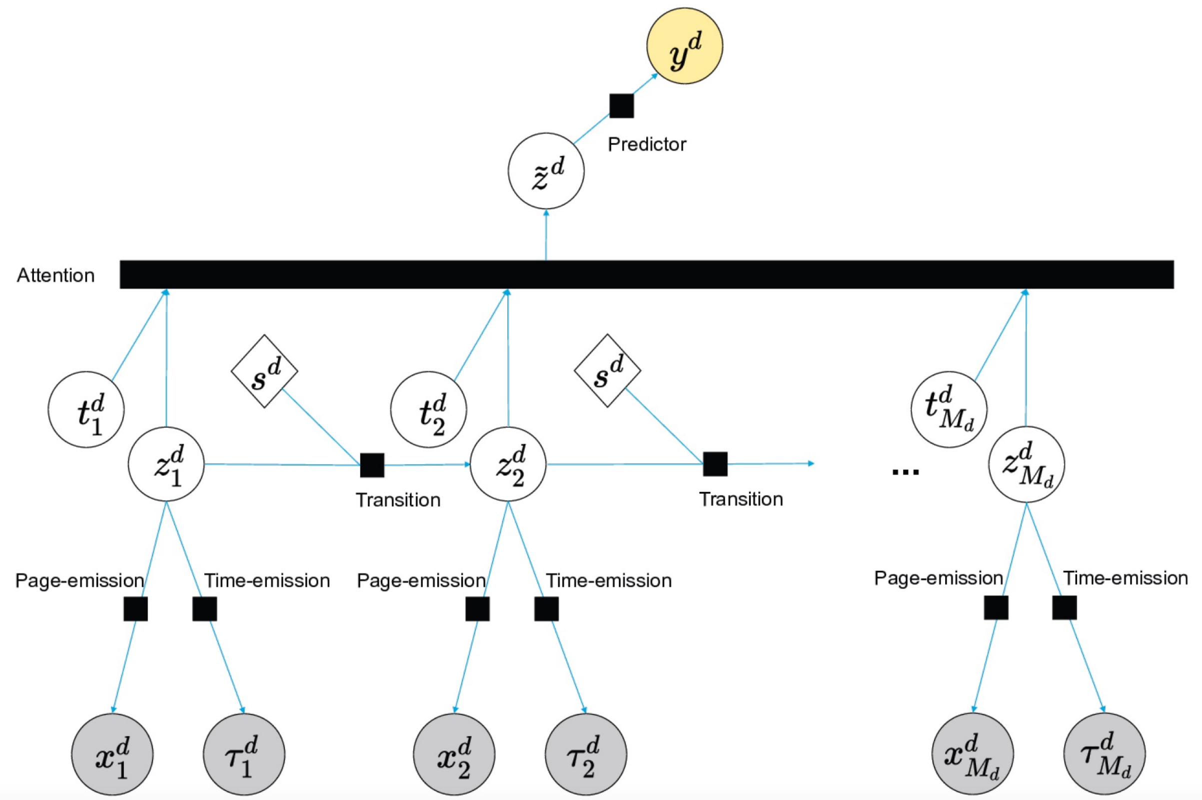
$$\tilde{z} = \sum_{m=1}^{M} \gamma_m\, z_m$$

- **Predictor network** is the final step of making the inference of purchase probability.

$$\hat{y} = \mathrm{sigmoid}(U_y\, \mathrm{ReLU}(W_y\, \tilde{z} + b_y) + c_y)$$

Model

Figure 1: Model Architecture of *Clickstream*DMM. Black squares denote neural networks.



Results

Purchase prediction with early warnings

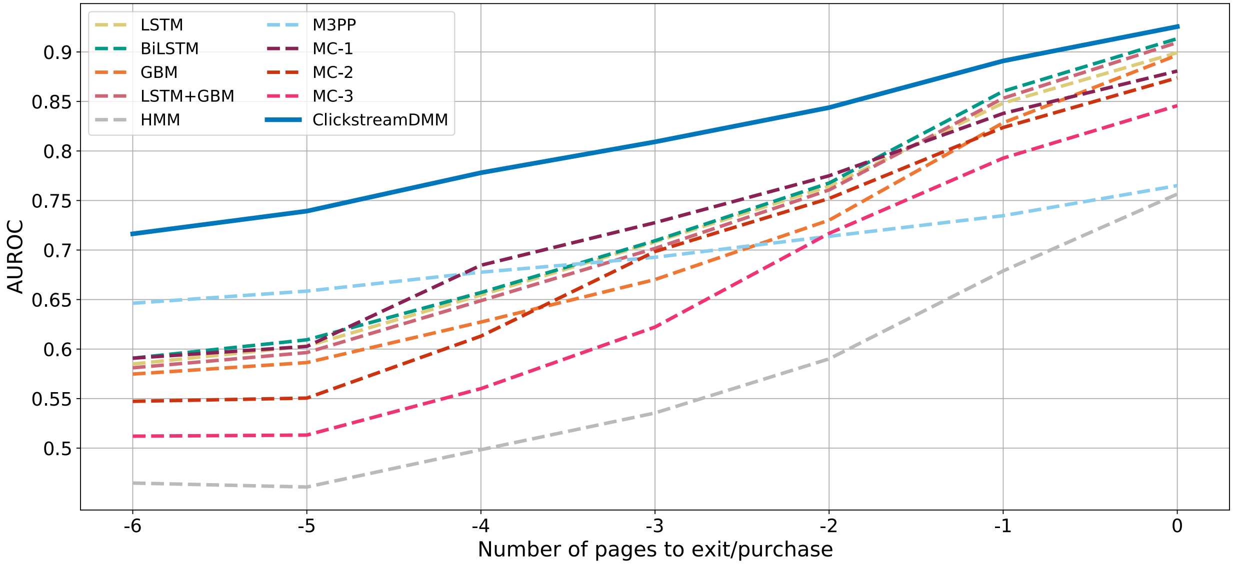
Table 2: The risk of exiting with no purchase is estimated starting from n = 6 pages ahead to exit

Model	AUROC	AUPRC
LSTM [16, 32]	0.726 ± 0.014	0.472 ± 0.012
BiLSTM [39]	0.733 ± 0.017	0.497 ± 0.015
GBM	0.705 ± 0.006	0.500 ± 0.011
LSTM+GBM [16]	0.725 ± 0.011	0.505 ± 0.012
HMM [5, 24]	0.572 ± 0.005	0.259 ± 0.003
M3PP [11]	0.701 ± 0.012	0.447 ± 0.012
MC-1 [18]	0.732 ± 0.006	0.484 ± 0.009
MC-2 [18]	0.697 ± 0.007	0.456 ± 0.010
MC-3 [18]	0.655 ± 0.009	0.407 ± 0.011
***Clickstream*DMM** (ours)	$\mathbf{0.817 \pm 0.007}$	$\mathbf{0.569 \pm 0.013}$

Higher is better. Best value in bold.

Illustration

Figure 2: Prediction performance when making predictions n = 6 pages ahead to exit



Conclusion

- **Theory informed model:** *Clickstream*DMM jointly captures **long-term dependence** of page clicks (via an attention network) and **different latent shopping phases** of users (via a latent variable model).

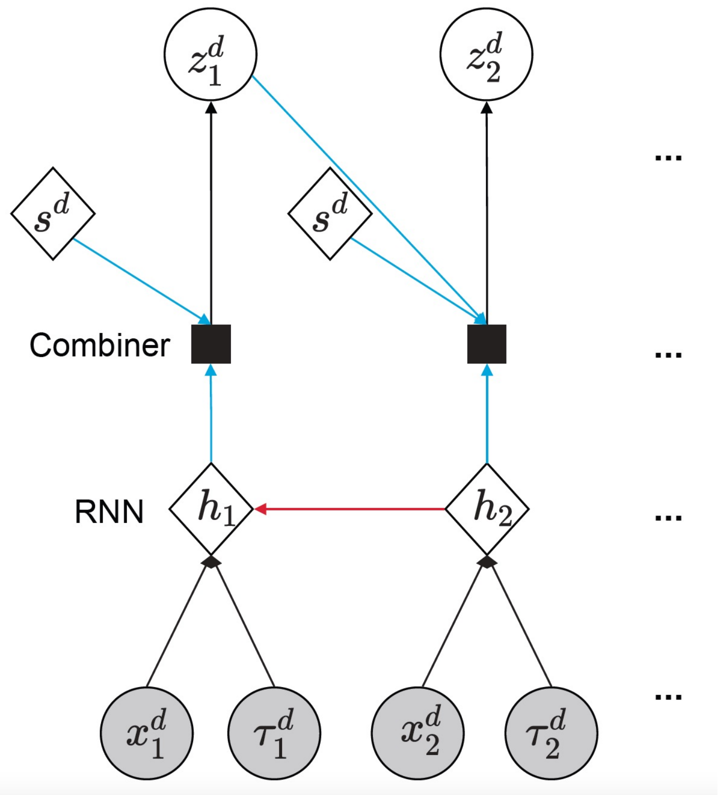
- **Performance**: *Clickstream*DMM is particularly effective when predictions are made multiple pages ahead. This is needed upon deployment, so that early warnings are generated to trigger marketing interventions.

- **Generalizability:** *Clickstream*DMM can be applied to other settings where time-series data has long-term dependencies and are driven by latent dynamics, such as healthcare analytics or churn prediction.

- Don’t forget to check our paper ***A Deep Markov Model for Clickstream Analytics in Online Shopping*** in the ACM Web Conference 2022 (**WWW ’22**).

Posterior Approximation

Figure 3: Posterior Approximation of *Clickstream*DMM. Black squares denote neural networks.



- Posterior distribution of the latent variables is approximated as

$$q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s,) = \prod_{m=1}^{M} q(z_m \mid z_{m-1}, x_{1:M}, \tau_{1:M}, s)$$

- The formula above is further simplified as

$$q(z_m \mid z_{m-1}, x_{1:M}, \tau_{1:M}, s) = q(z_m \mid z_{m-1}, x_{m:M}, \tau_{m:M}, s)$$

Loss Function

- *ClickstreamDMM* minimizes the following loss

$$\mathcal{L}(y, \hat{y}, \mathcal{S}) = \ell(y, \hat{y}) - \alpha\, \mathrm{ELBO}(\mathcal{S})$$

as a combination of

- **Cross-entropy loss**

$$\ell(y, \hat{y}) = -\rho\, y \,\log\,(\hat{y}) - (1 - y)\,\log\,(1 - \hat{y})$$

- **Evidence Lower Bound (ELBO)**

$$\mathrm{ELBO}(\mathcal{S}) = \mathbb{E}_{q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s)} [\log p(x_{1:T}, \tau_{1:T} \mid z_{1:T})]$$
$$- \mathrm{KL}(q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s) \,||\, p(z_{1:M})).$$

Visualization of Latent Variables

Figure 4: Clustering of latent variables

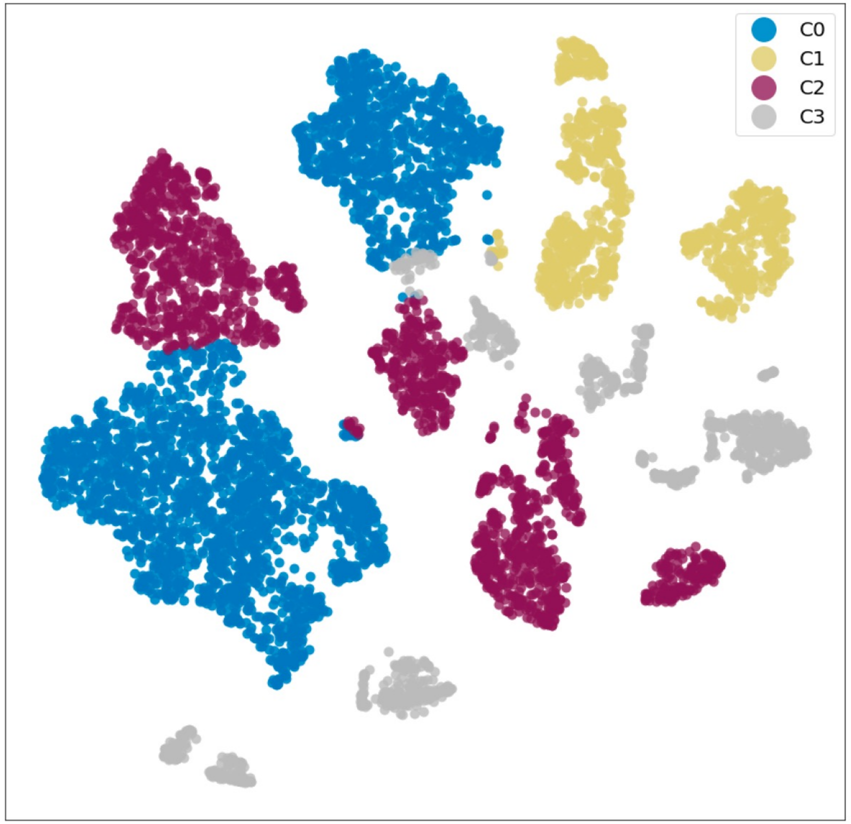
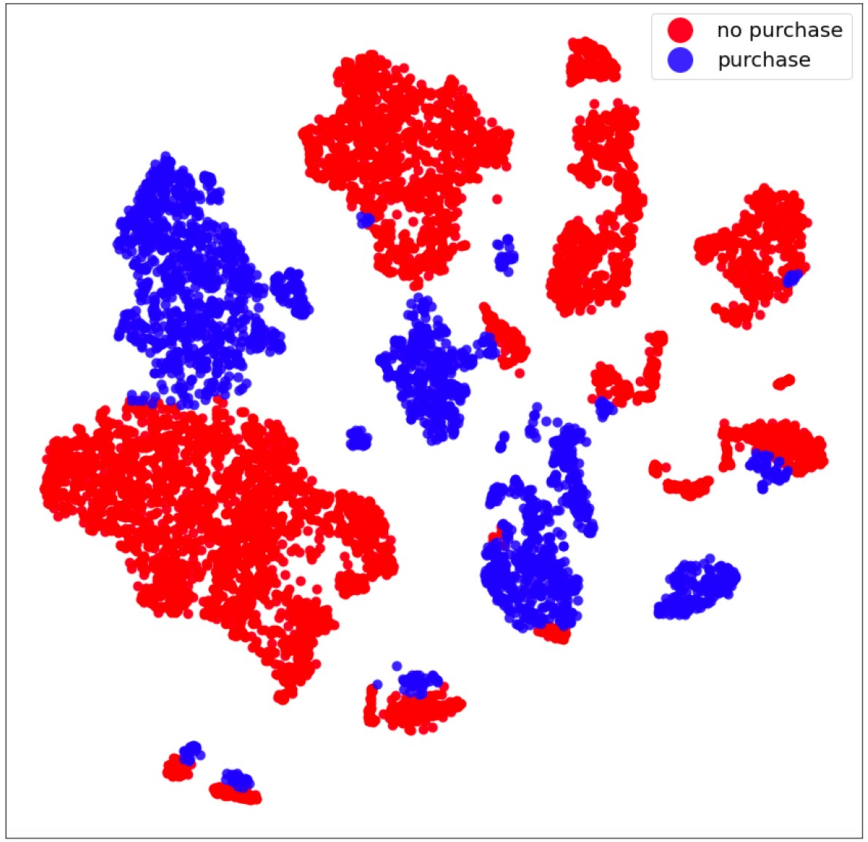


Figure 5: Latent variables of sessions with and without purchase



Characteristics of Clusters

Figure 6: Relative frequency of page clicks by cluster (in %)

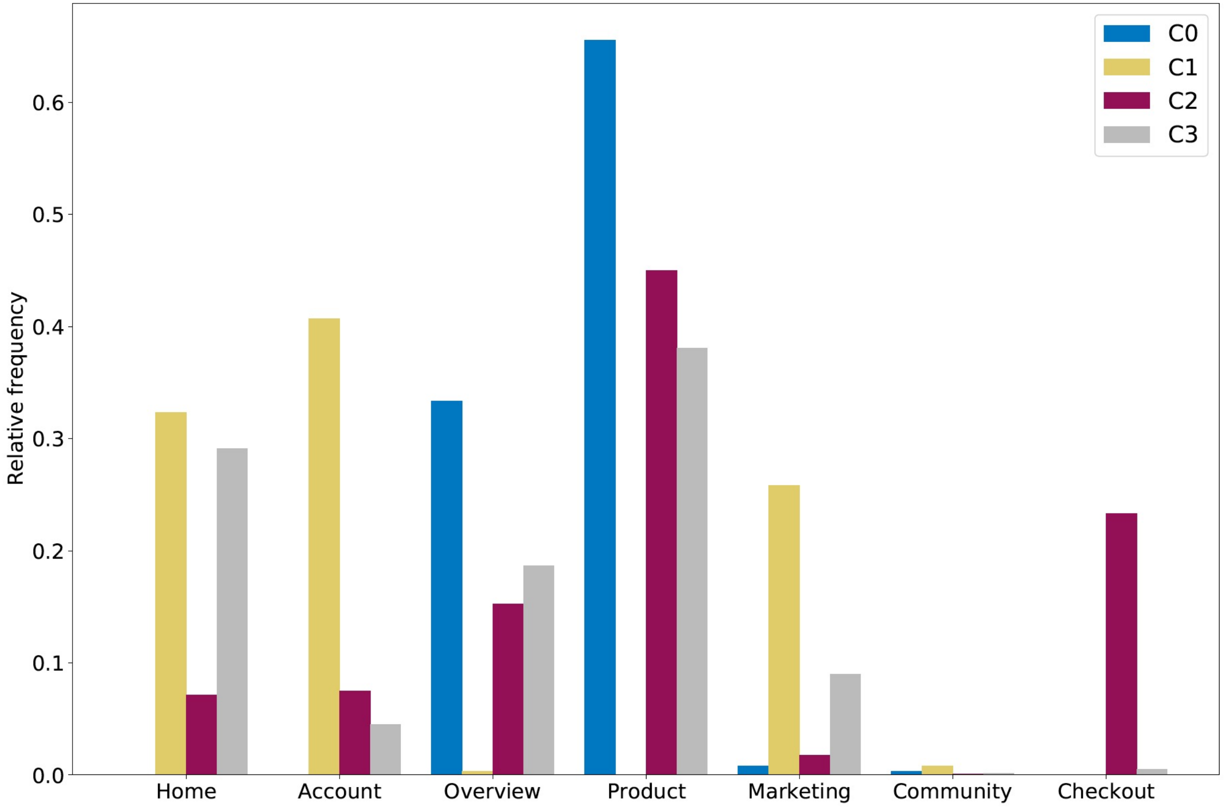
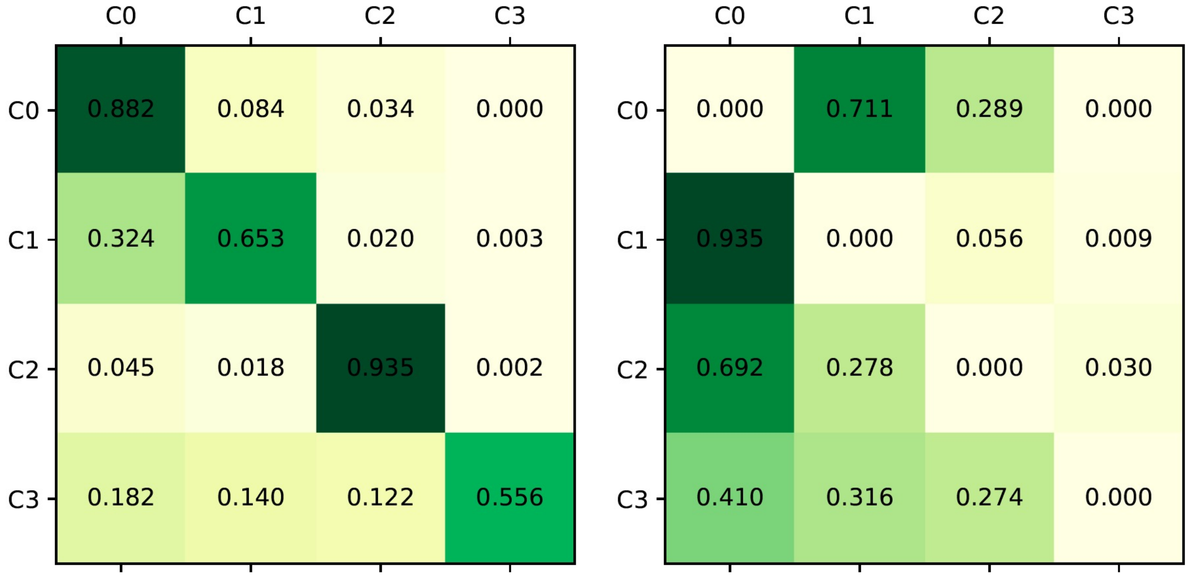


Figure 7: Transition matrix between clusters (left: current, top: next)



a) with self-transition

b) without self-transition