TEAM 9

Pretrained CoLES model quality based on the input data amount

Anastasia Volkova
Olga Volkova
Ksenia Kuvshinova
Alexander Zubrey
Anastasia Grigoreva

Skoltech

Content

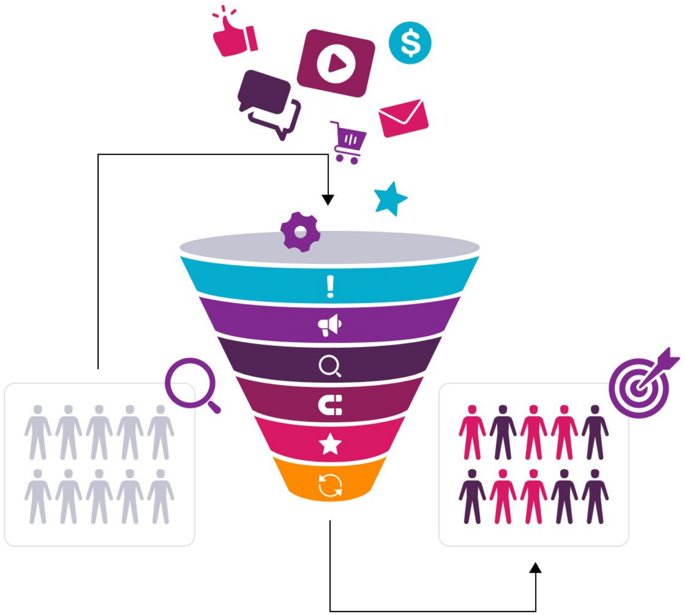
Motivation

Sequential data embeddings

Using of Pre-training data

Performance metrics

NN saturation



Problem Statement

	CoLES	Our project
Challenges	difficulty capturing temporal dependencies and patterns in customer transactions	CoLES does not have dependence of the model quality on the size of the pretrain data

Problem Statement

	CoLES	Our project
Challenges	difficulty capturing temporal dependencies and patterns in customer transactions	CoLES does not have dependence of the model quality on the size of the pretrain data
Solution	a new self-monitoring method for embedding discrete event sequences based on contrastive learning	different sizes of pretrain testing, NN saturation investigation, comparison of model accuracy

Description of CoLES metod

Algorithm for random slices sub-sequence generation strategy

Input: the sequence of length T $S = \{z_j\}_{j=0}^{T-1}$

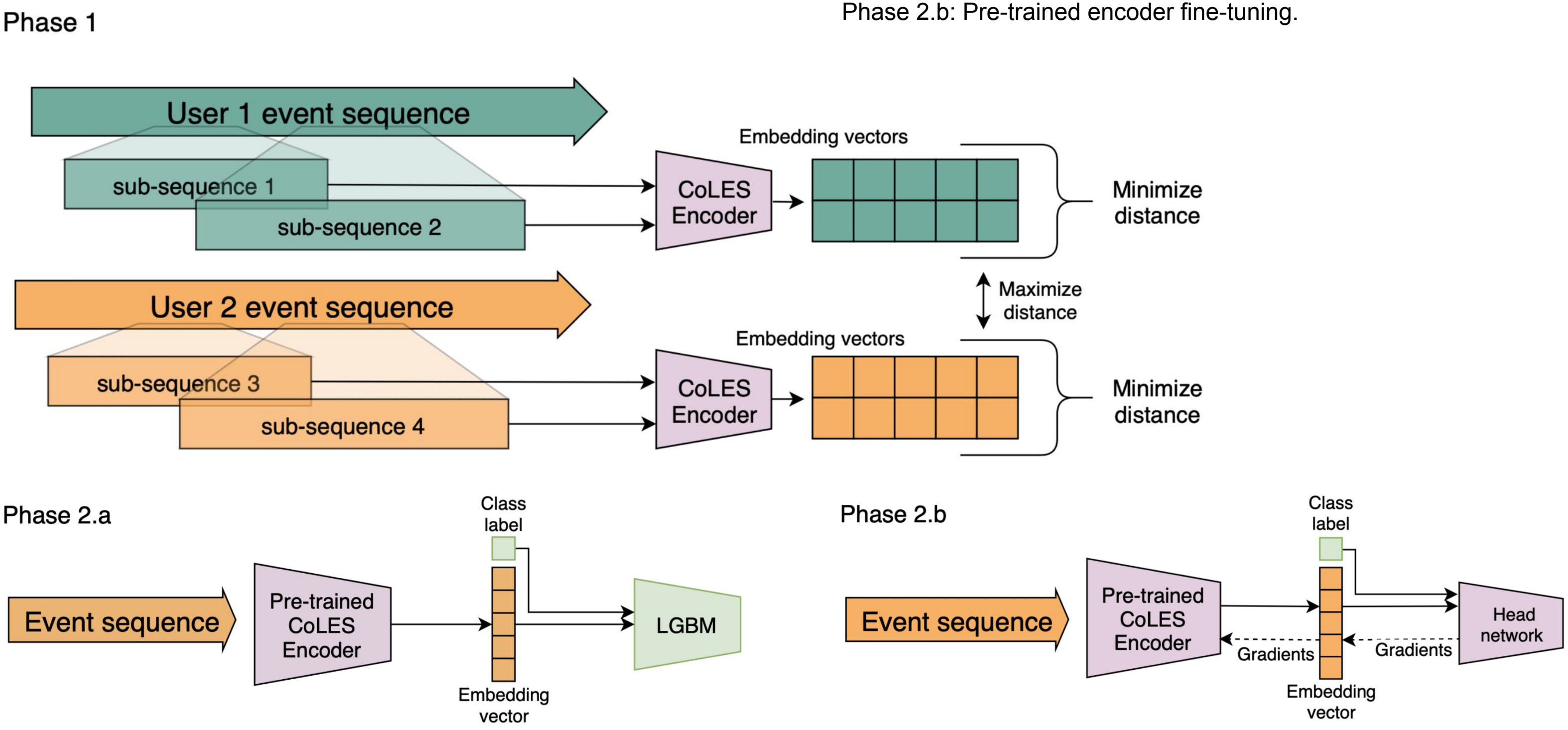
Output: $\mathcal{S}$ subsequence of S

for $i \leftarrow 1$ **to** k **do**
- Generate a random integer T_i uniformly from $[1, T]$;
- **if** $T_i \in [m, M]$ **then**
  - Generate a random integer s from $[0, T - T_i)$;
  - Add the slice $\tilde{S}_i := \{z_{s+j}\}_{j=0}^{T_i-1}$ to $\mathcal{S}$;

end

Description of CoLES metod

Figure 1: General framework. Phase 1: Self-supervised training.
Phase 2.a Self-supervised embeddings as features for supevised model.
Phase 2.b: Pre-trained encoder fine-tuning.



Dmitrii Babaev, Nikita Ovsov, Ivan Kireev, Maria Ivanova, Gleb Gusev, Ivan Nazarov, and Alexander Tuzhilin. 2022. CoLES: Contrastive Learning for Event Sequences with Self-Supervision. In Proceedings of the 2022 International Conference on Management of Data (Philadelphia, PA, USA) (SIGMOD'22). Association for Computing Machinery, New York, NY, USA, 1190–119

Description of CoLES metod

CoLES

Event Sequent Encoder

Pairs Generation Strategy

Loss Function for Contrastive Algorithm

$$\mathcal{L}_{uv}(M) = Y_{uv}\frac{1}{2}d_M(u,v)^2 + (1 - Y_{uv})\frac{1}{2}\max\{0, \rho - d_M(u,v)\}^2 \quad \text{wrt} \quad M: \mathcal{X} \to \mathbb{R}^n$$

$d_M(u,v) = d(c_u, c_v)$ - distance between embeddings of the pair (u,v)

$c_* = M(\{\tilde{x}_*(\tau)\})$ ρ - soft minimal margin between dissimilar objects

$Y_{uv} = 1$ - if samples from same sequences $Y_{uv} = 0$ - if samples from different sequences

Experimental Setup

Transactions Data

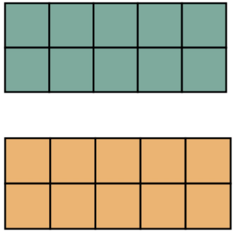
росбанк

Binary

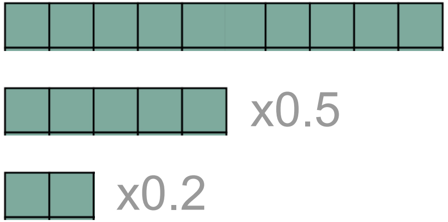
СБЕР

MultiClass

Embeddings coding by CoLES



Pretrained Data

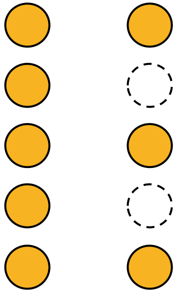


x1

x0.5

x0.2

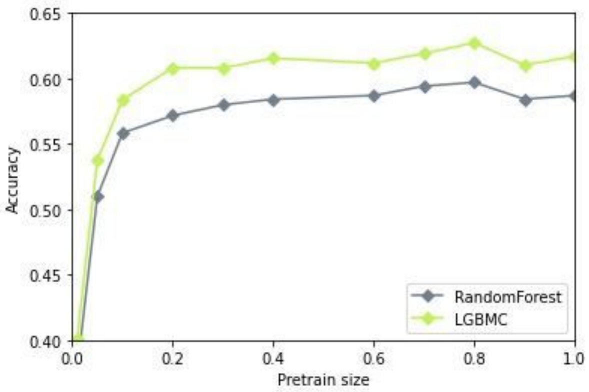
Model Size

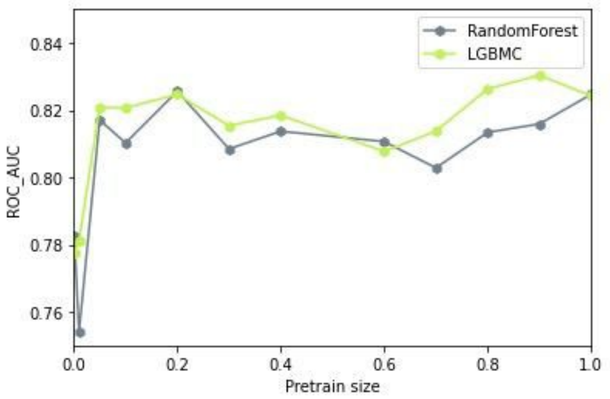


Experiments

СБЕР

росбанк





Accuracy

Hidden size	RandForest	LGBM
64	0.569	0.603
128	**0.592**	0.617
256	0.587	0.612
512	0.584	**0.623**

AUC_ROC

Hidden size	RandForest	LGBM
64	0.799	0.814
128	0.807	0.823
256	0.811	0.808
512	**0.816**	**0.834**

Conclusions

- CoLES method for building embeddings of discrete event sequences was implemented
- Pretrained dataset size on both transactions datasets had a strong influence on model performance. The “plato” can be observed at 40% of pretrained data
- Saturation of NN in self-supervised training was reached at half of value of the original model

Thank You!

Anastasia Volkova
Olga Volkova
Ksenia Kuvshinova
Alexander Zubrey
Anastasia Grigoreva

TEAM 9

Skoltech