Skoltech

Self-supervised learning for Event Sequences on synthetic task of next item prediction

27 club:

Egor Fadeev

Alexander Ganibaev

Matvey Lukyanov

Aleksandr Yugay

Outline

1. Self-supervised learning: motivation
2. Datasets
3. Methodology
4. Preliminary results

Self-supervised learning

- **Motivation:** learning the structure of the unlabeled data through supervised methods
- **Approach:**
  a) Features are learned using unlabeled data
  b) Object-target pairs are constructed from data points
  c) These pairs are used for learning the structure of the data

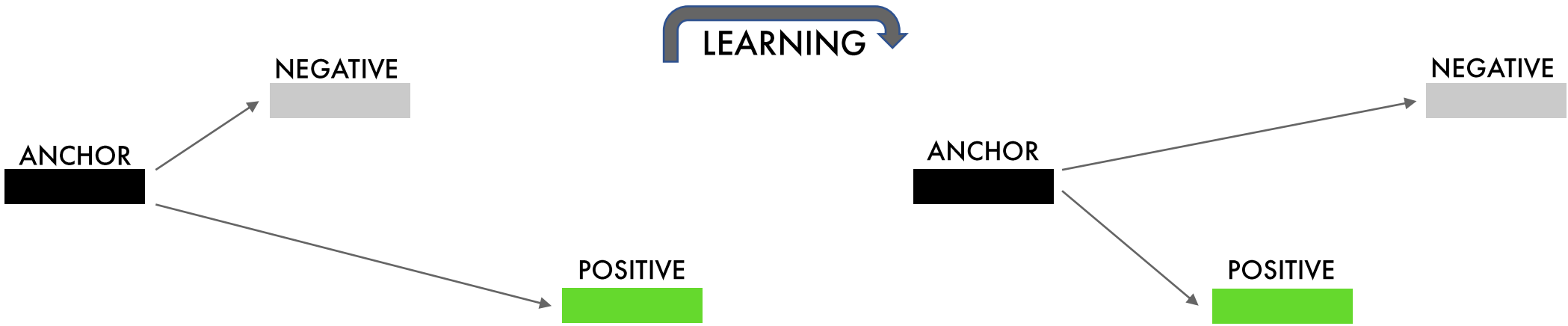
Can be used in various tasks, e.g. **representation learning**

Representation learning

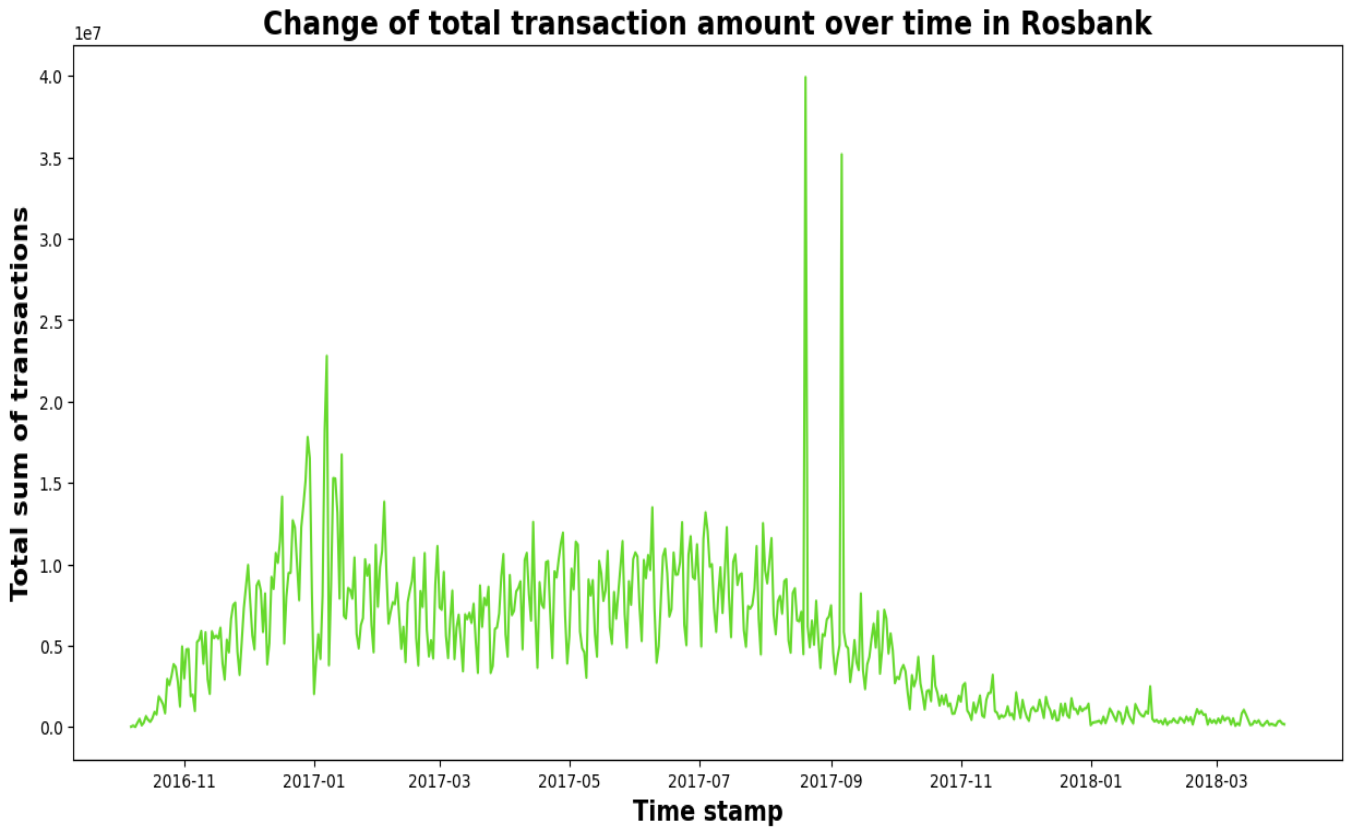
- **Motivation:** ML tasks often require input that is convenient to process
- **Obstacle**: for real-world data, e.g. transactions, specific features cannot be defined algorithmically
- **Solution:** representing unstructured data as the vectors of fixed-length
- **Usage:** embeddings of bank transactions for fraud detection, risk assessment, etc.

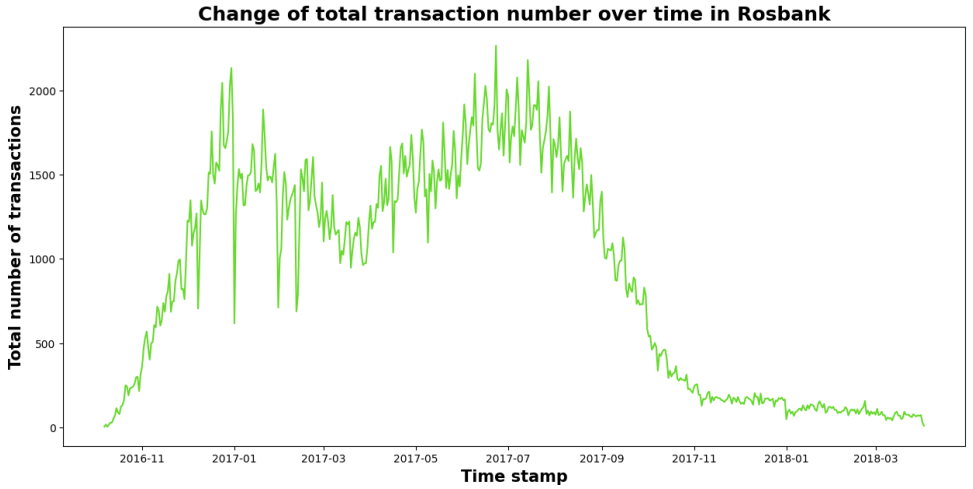
Contrastive learning

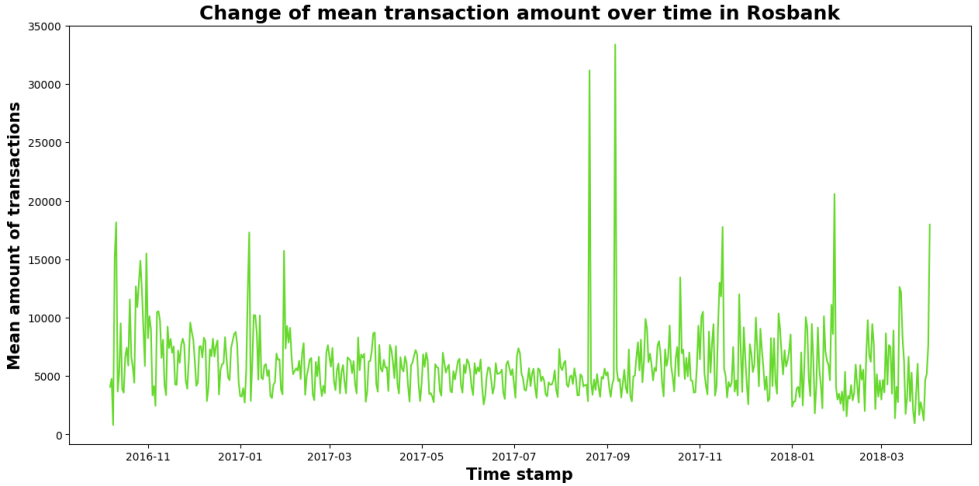
- **Motivation:** to prevent mapping the objects to the same representation
- **Approach:** involves training a model to differentiate between similar and dissimilar examples
- **Usage:** learn representations that differentiate between fraudulent and non-fraudulent transactions
- **Examples:** CoLES, CPC



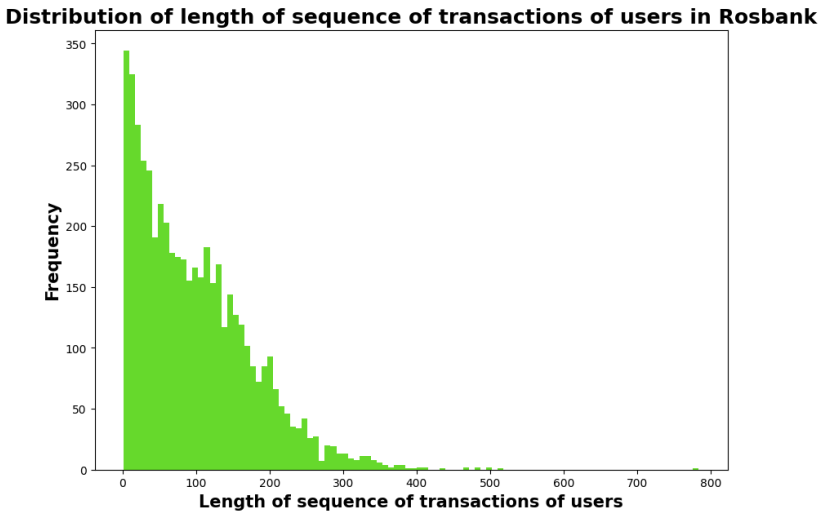
Datasets: rosbank

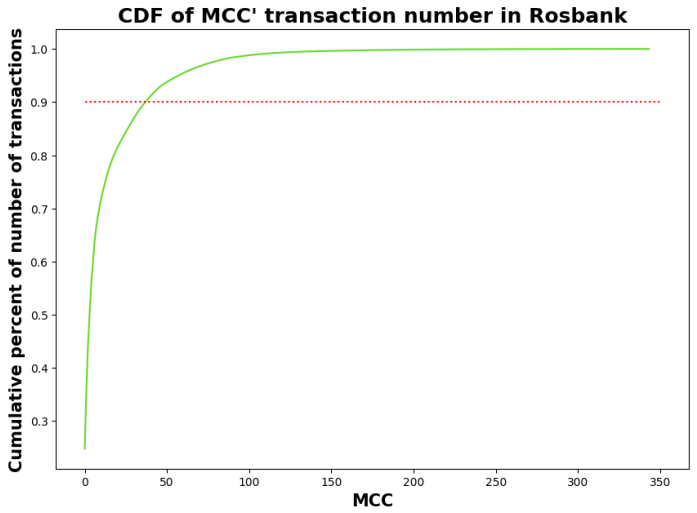


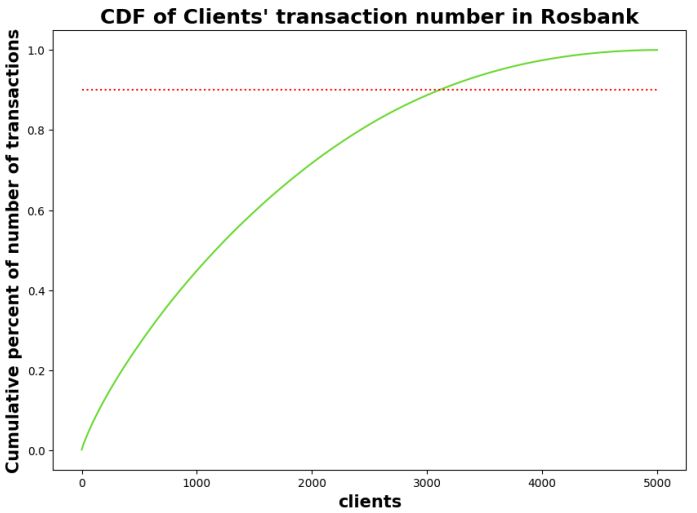




Datasets: rosbank

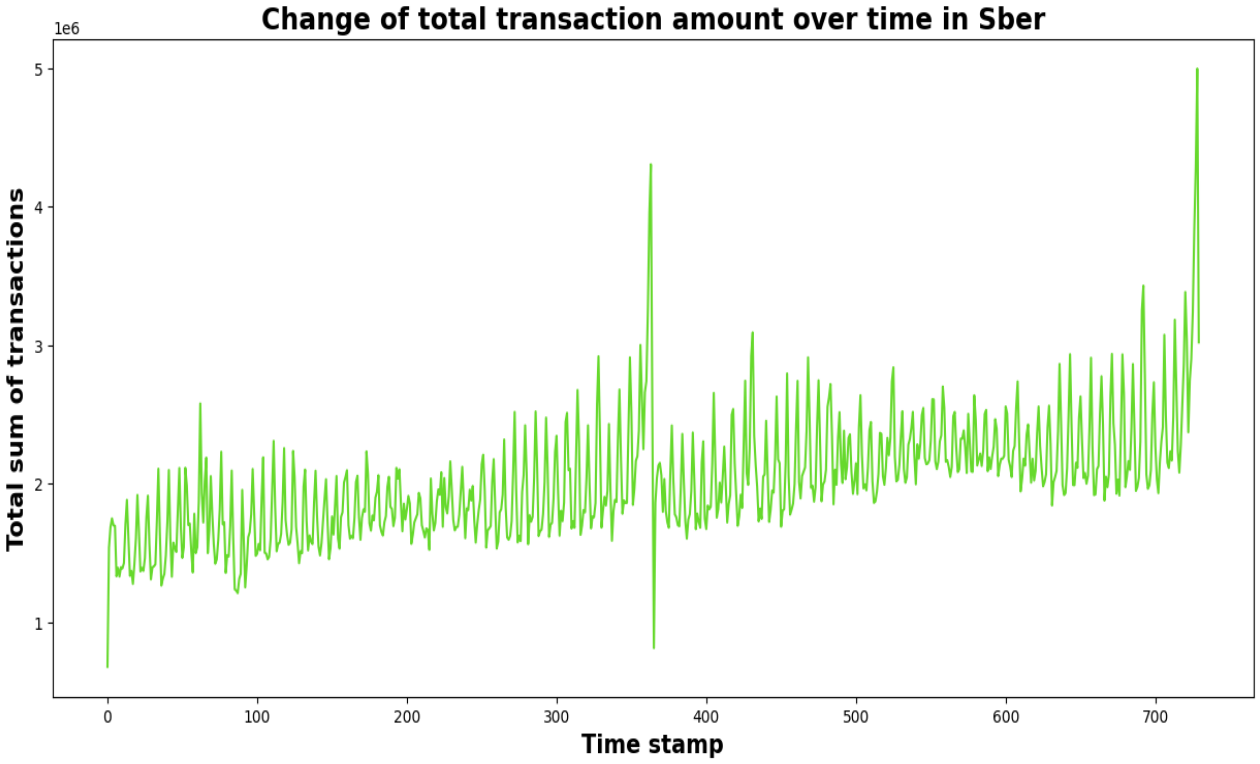


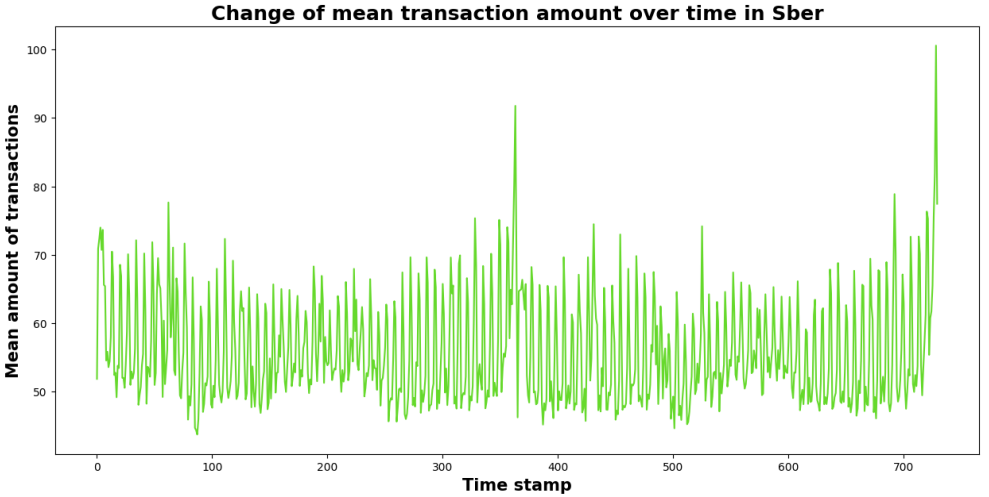


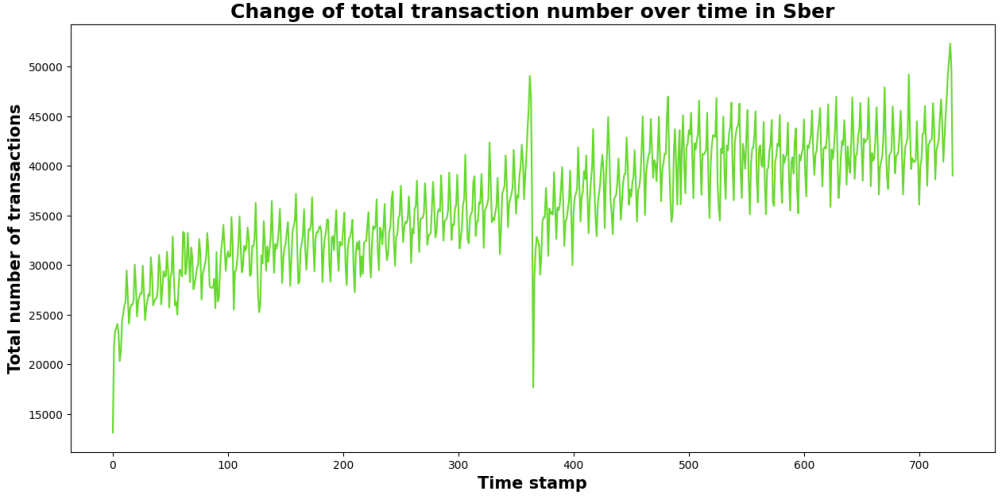


- **Problem:**
  - many clients with small number of transaction
  - many MCCs with small number of transactions
- **Possible solutions:** drop such clients and unite such MCCs in ‘other’ group

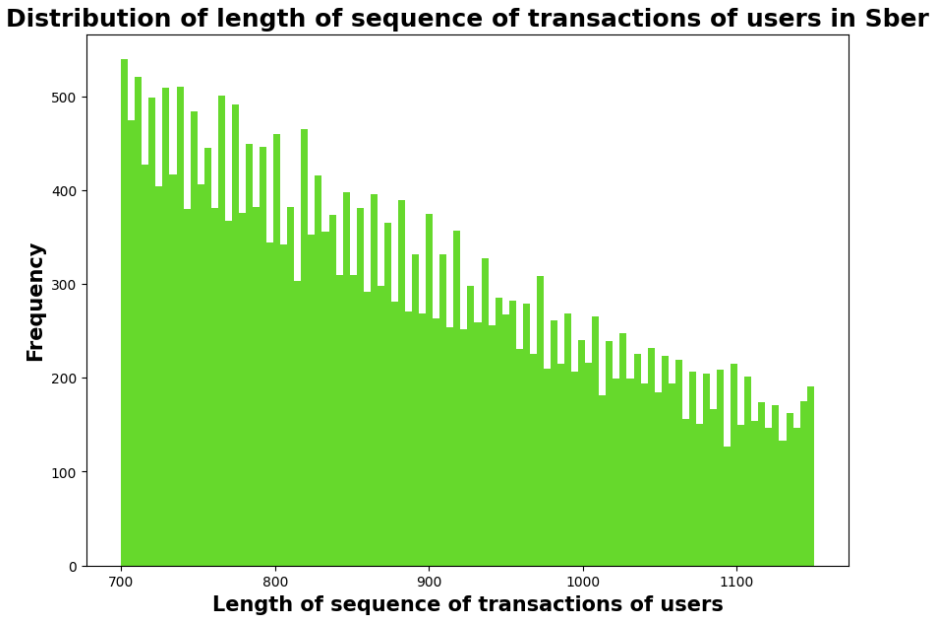
Datasets: sberbank

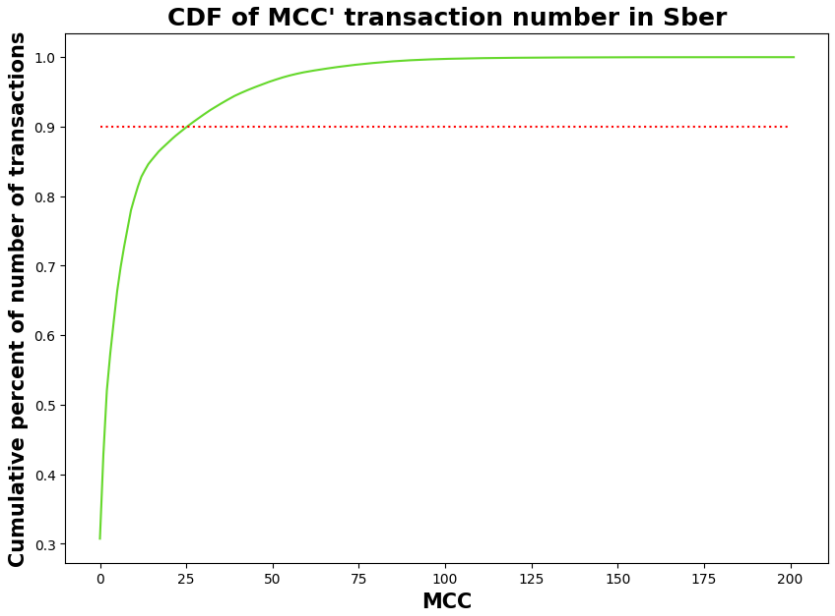






Datasets: sberbank

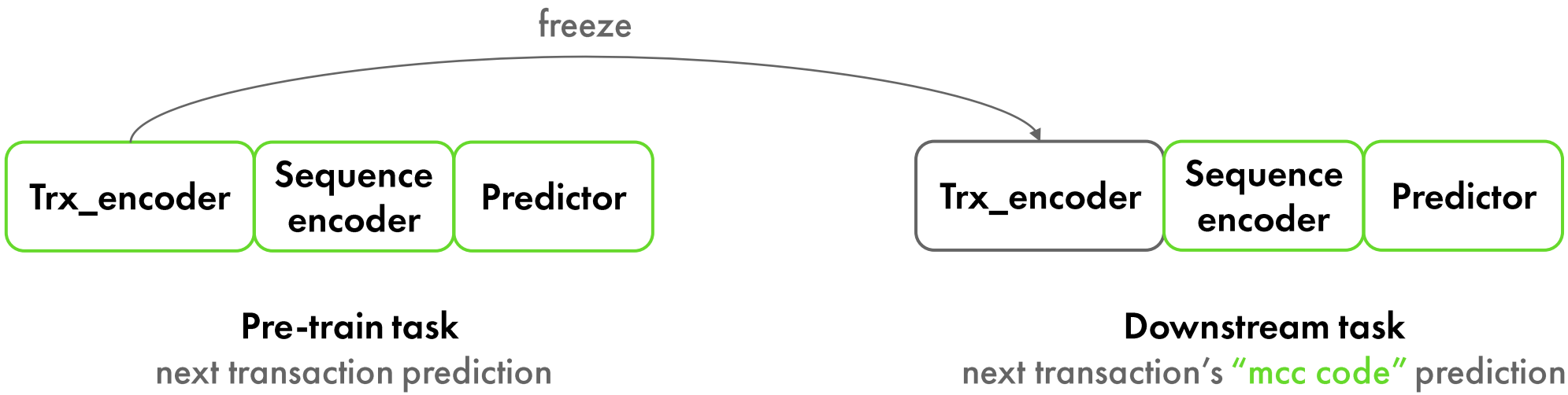




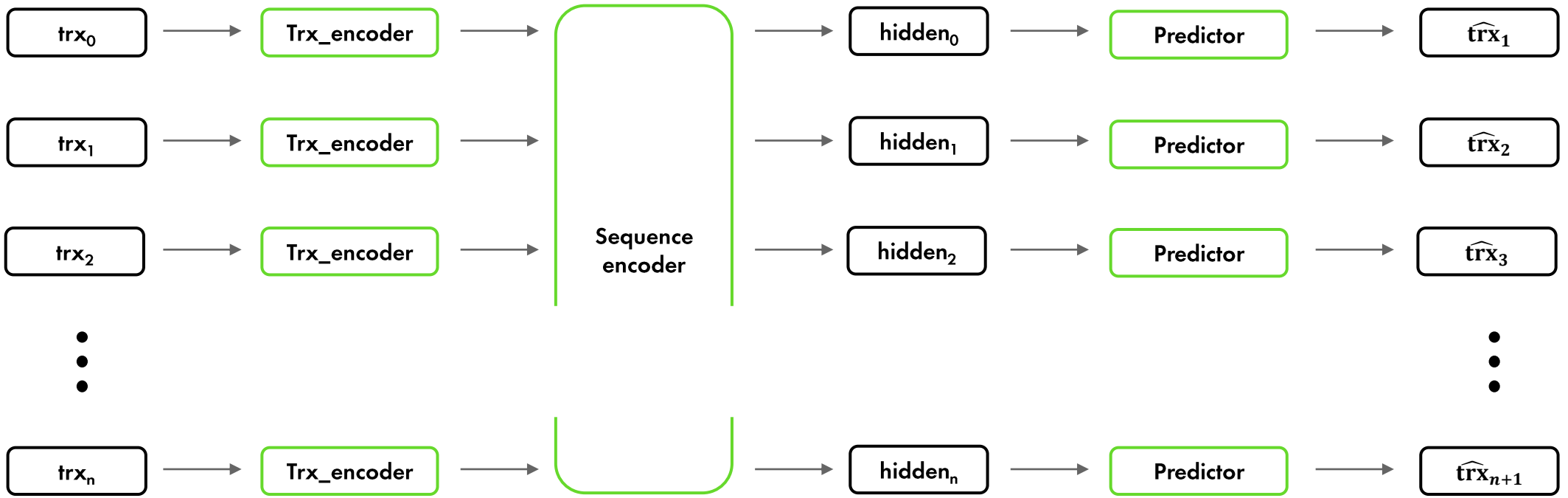
- **Problem:** many MCCs with small number of transactions
- **Solution:** drop such MCCs

Methodology

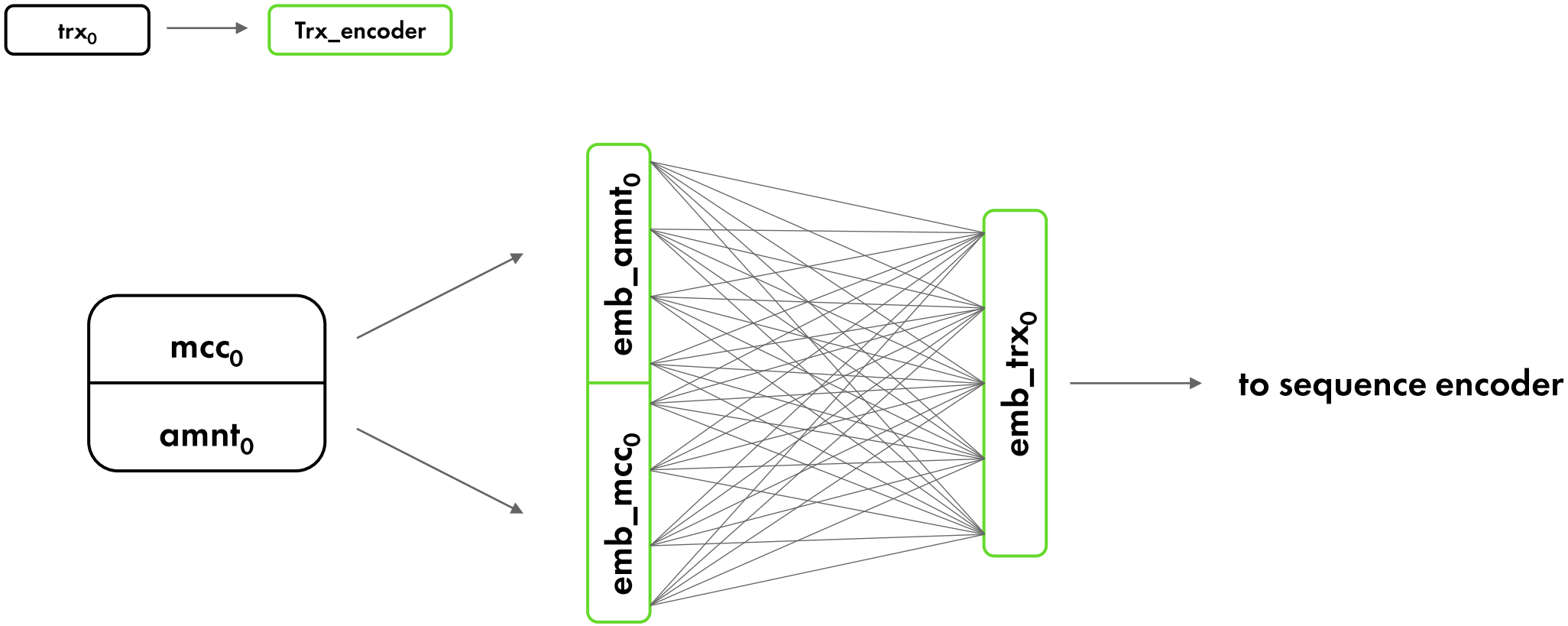
- **Features:** "mcc code" and "amount"
- **Data preprocessing:** discretization of "amount" feature (10 bins)
- **Train/validation/test split:** 80%/10%/10%
- **Metric**: f1-score weighted



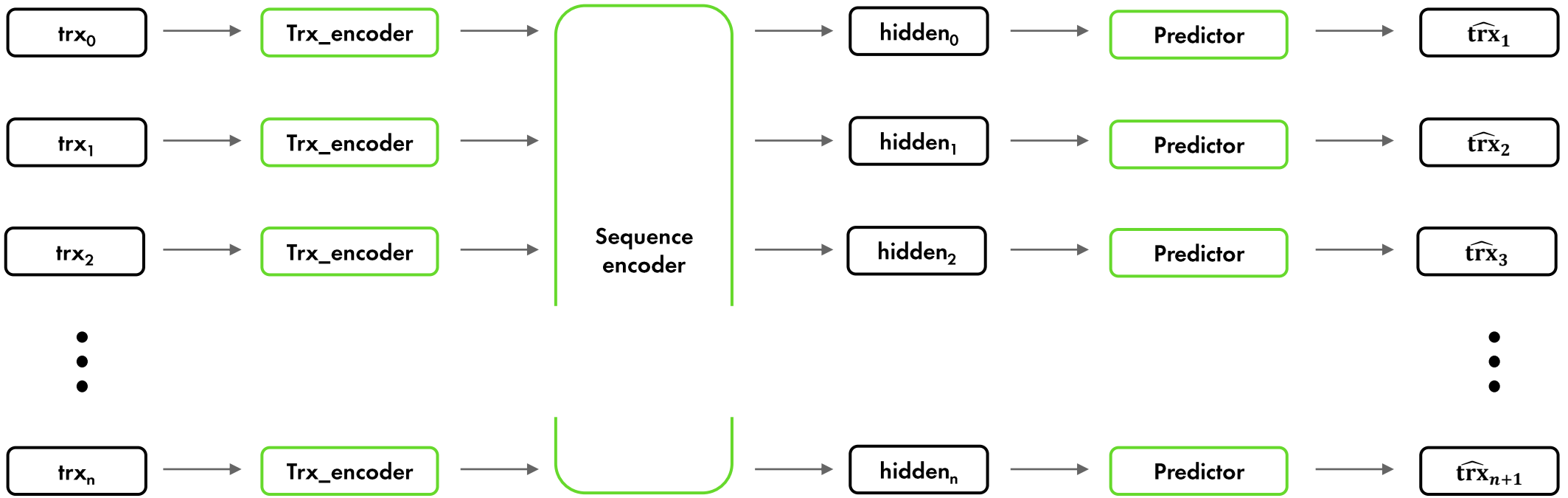
Representation (1)



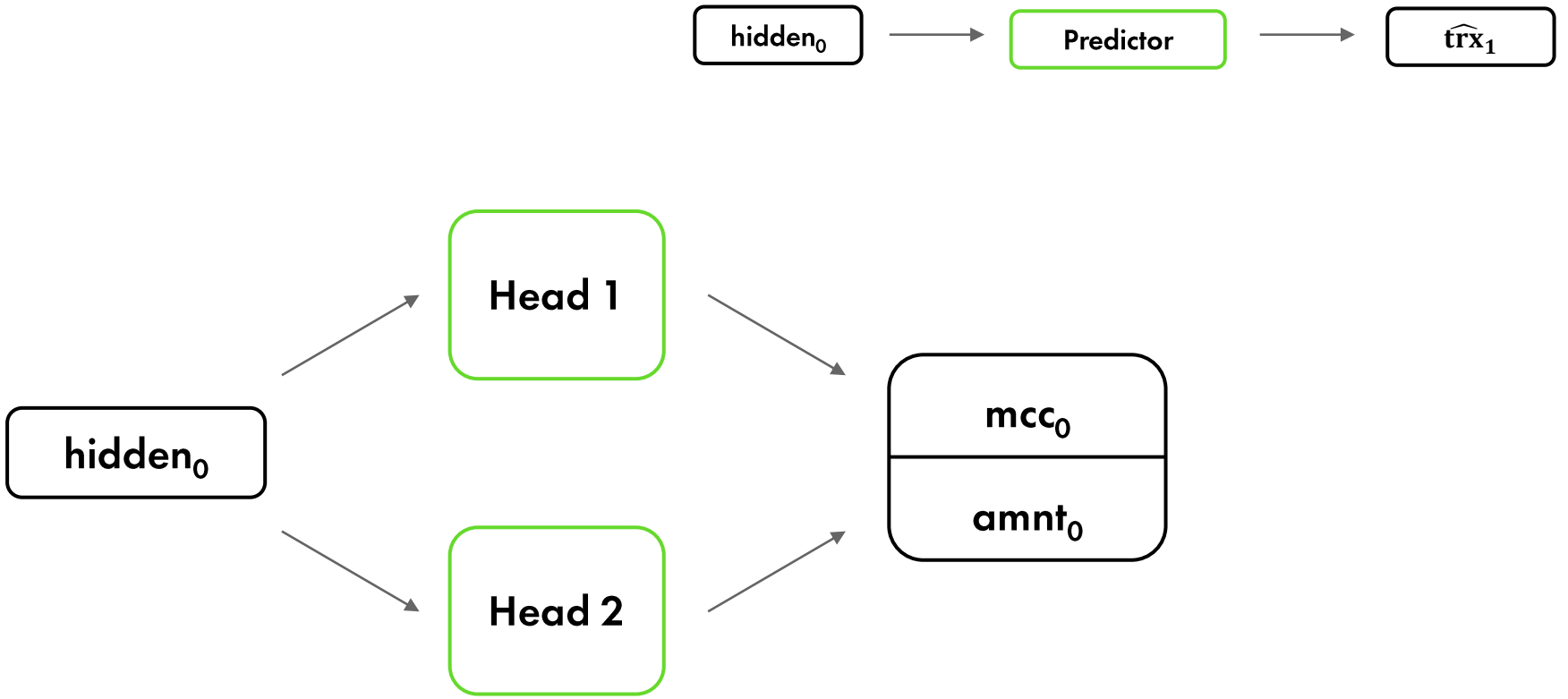
Representation (2)



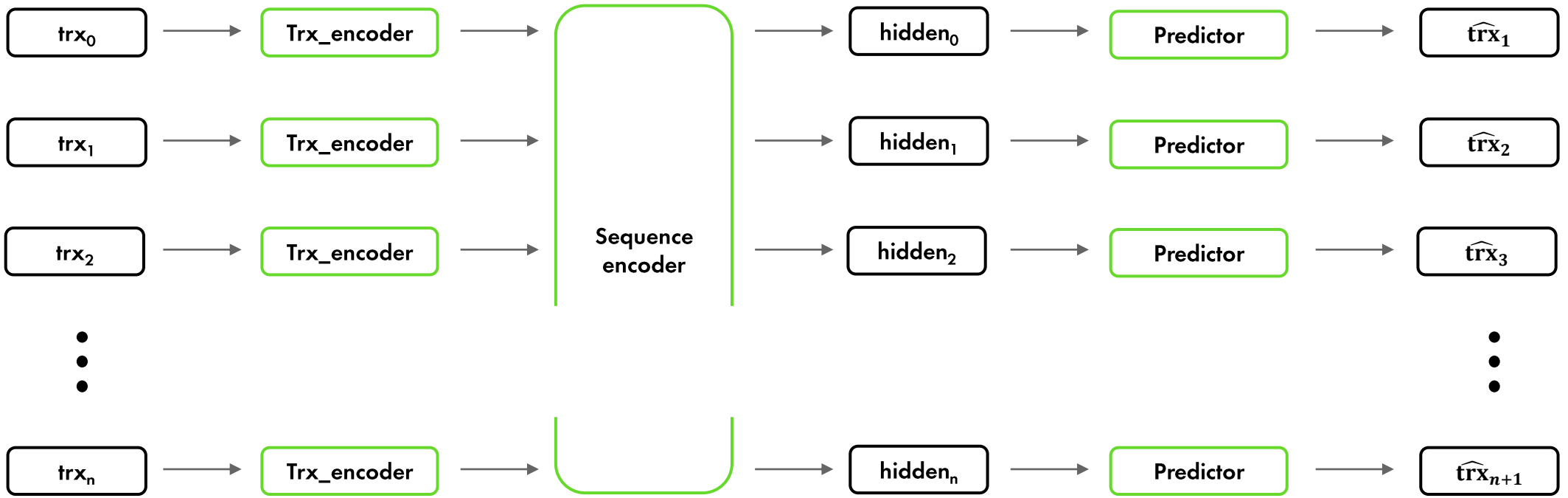
Representation (3)



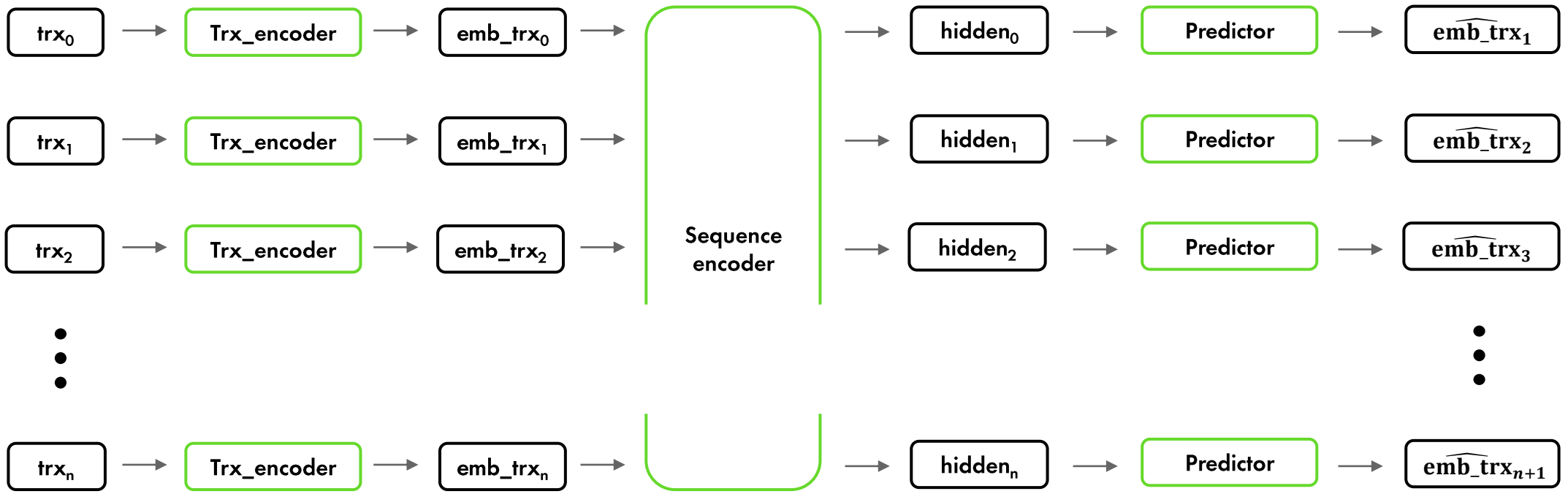
Representation (4)



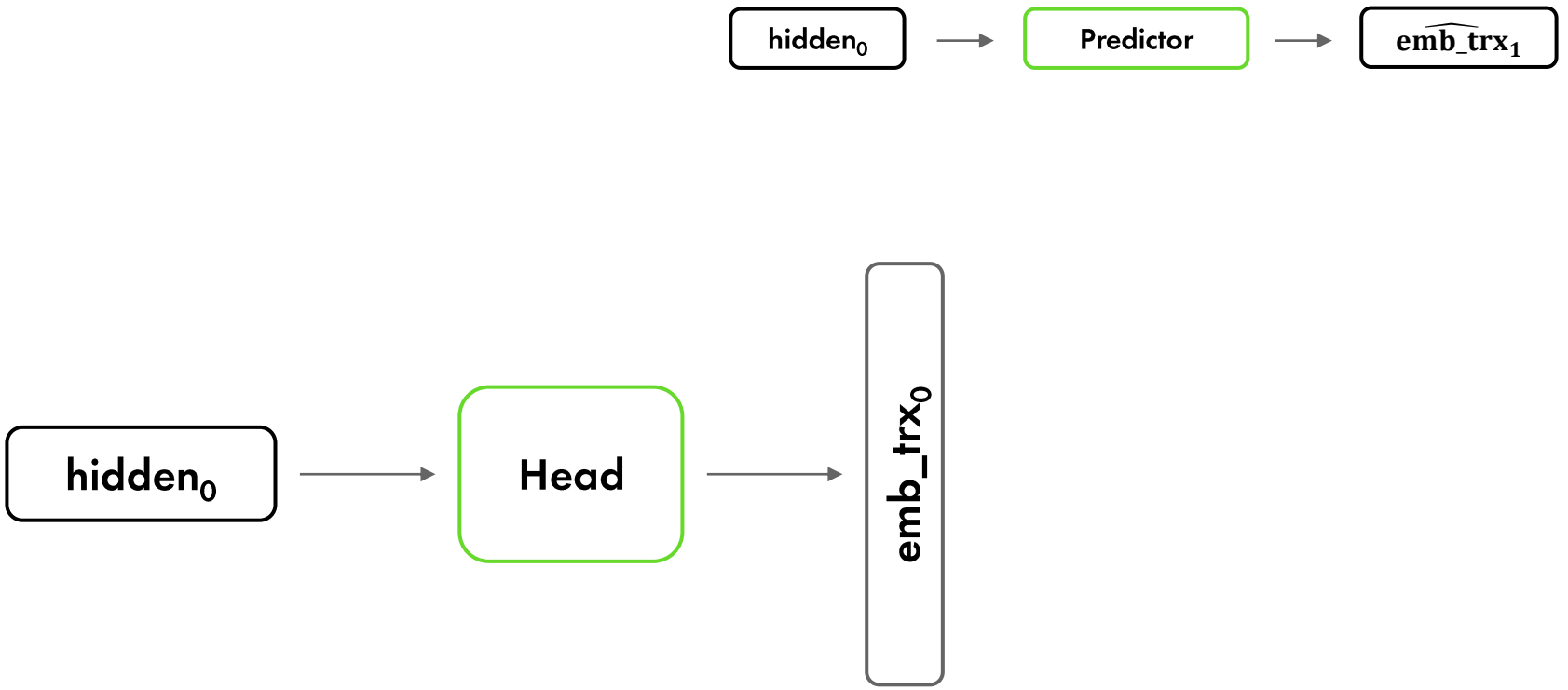
Representation (5)



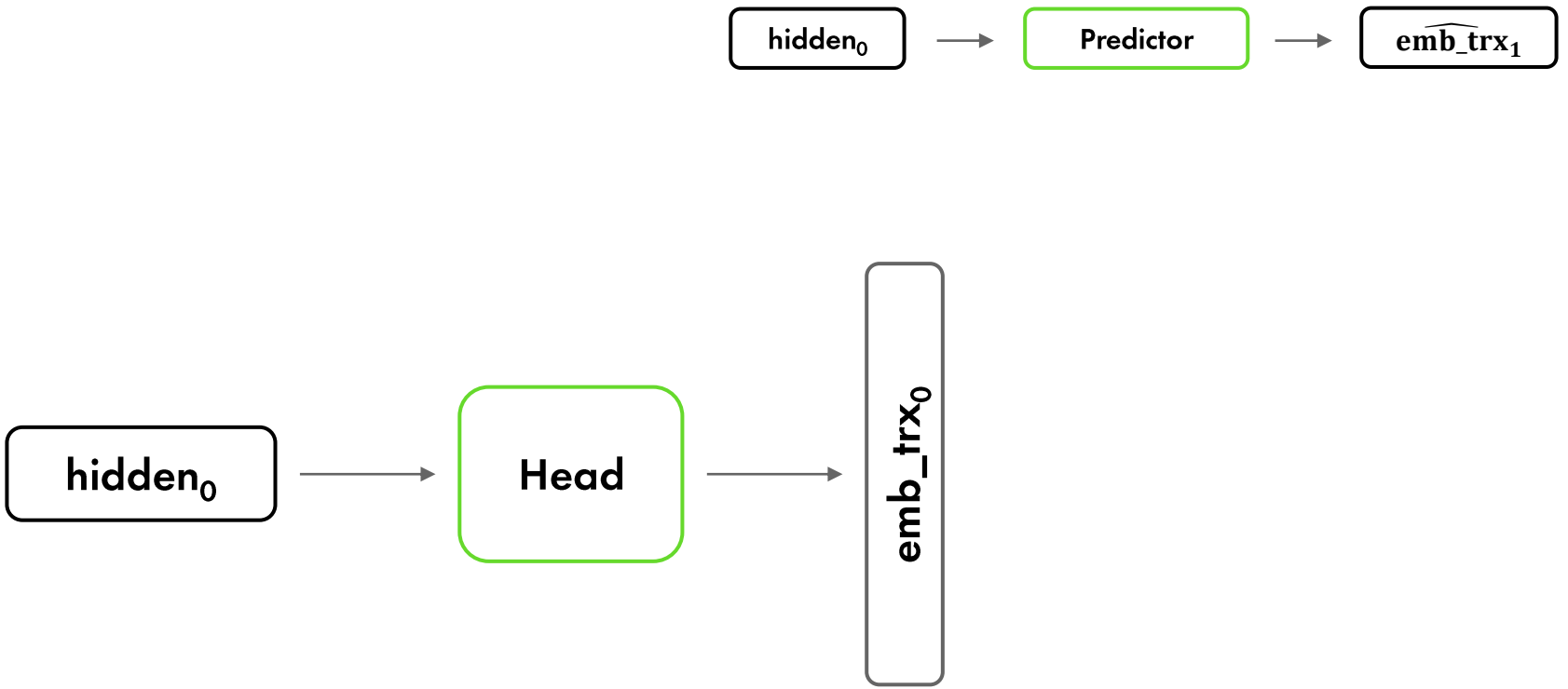
Contrastive (1)



Contrastive (2)

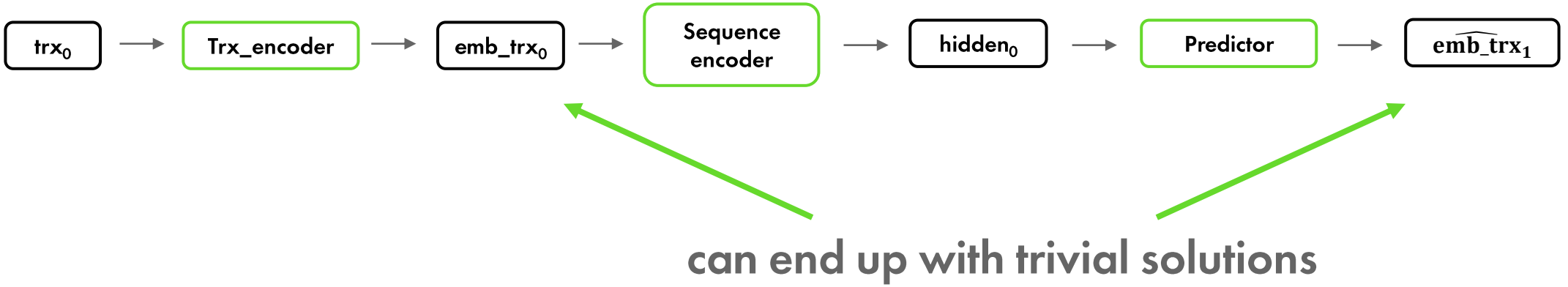


Contrastive (2)

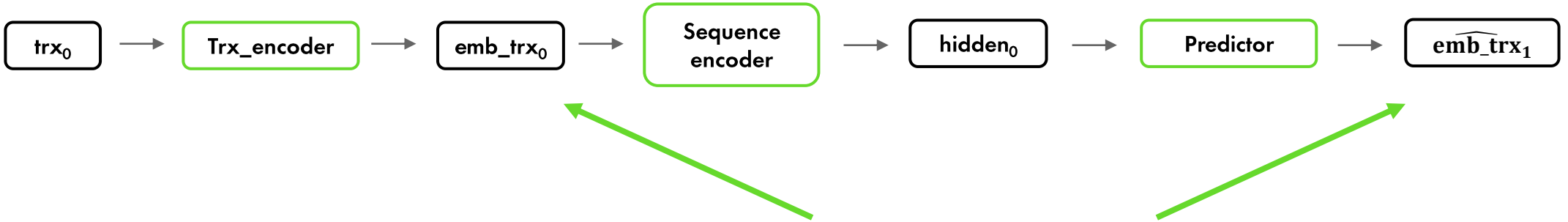


Why is it called contrastive?

Contrastive (3)



Contrastive (3)



can end up with trivial solutions

QuerySoftmaxLoss

Preliminary results

- Tests with the same hyperparameters were ran
- No significance difference in Rosbank dataset for 20 runs

	Representation	Contrastive
Mean	0.23468	0.23238
Std	0.00614	0.00754
Sample size	20	20