СБЕР

Multimodal banking data and event sequences

Ivan Kireev
Sber AI Lab

26.10.2024

Ivan Kireev

7 years in Machine Learning

5 years in Deep Learning (Sber AI Lab)

2 years as Head of Deep Learning Center

7 scientific publications

Website
Sber AI Lab

Research Interests:

- event sequences
- representation learning
- Matching
- LLM
- adversarial methods
- dynamic graphs

Deep Learning Center

Event Sequences

Neural Network Potential:

- Large data volume
- Complex data structure

Open Source Datasets

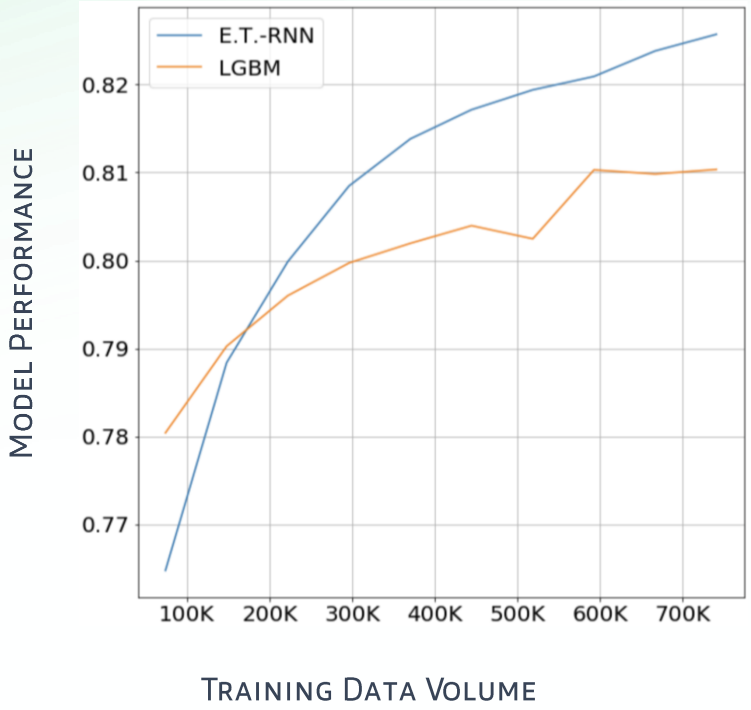
Sequence type	Target	Dataset
Financial history	Client's Gender and Age	SberAgePred
		SberGender
	Churn Prediction	Rosbank
		DataFusion
	Default Indicator	AlphaBattle
Retail Receipt History	Uplift	X5RetailHero
Learning App Logs	Exam Score	Bowl2019
URL Visit History	Client's Gender and Age	MtsMLCup
Music Listening Logs	Genre	YandexMLCup

Sequential Data – Example

cl_id	MCC	channel_type	currency	TRDATETIME	amount	trx_category
int64	int64	string · *classes*	int64	string · *lengths*	float64	string · *classes*
2 10.2k	742 9.41k	5 values	32 986	16 16	0.01 18.3M	10 values
3	6,011	null	810	09AUG17:20:08:44	2,000	WD_ATM_ROS
3	5,814	null	810	06JUL17:00:00:00	695	POS
3	5,999	null	810	21JUL17:11:20:12	100	POS
3	5,912	null	810	01JUL17:00:00:00	1,966	POS
3	5,411	null	810	27JUL17:00:00:00	360	POS
3	5,977	null	810	12JUN17:00:00:00	1,064	POS
3	6,011	null	810	14JUL17:00:00:00	5,500	WD_ATM_OTHER
3	5,814	null	810	12JUN17:00:00:00	187	POS
4	5,541	null	810	04FEB18:00:00:00	304	POS
4	6,012	null	810	21FEB18:13:03:19	700	C2C_OUT
4	5,631	null	810	18FEB18:00:00:00	373.5	POS
4	5,921	null	810	23MAR18:00:00:00	212.98	POS
4	5,814	null	810	09MAR18:00:00:00	622	POS

Creating a Universal Embedding on a Large Volume of
Unlabeled Data

Example: End-to-end training for scoring tasks
ET-RNN outperforms boosting in quality as the training data volume increases



Problem:

Neural networks require large amounts of labeled data for training, which are not always available for specific tasks.

Solution:

- Train a universal model using large volumes of unlabeled data.
- Adapt this model for individual tasks.

Technology:

Embedding as the output of a universal model

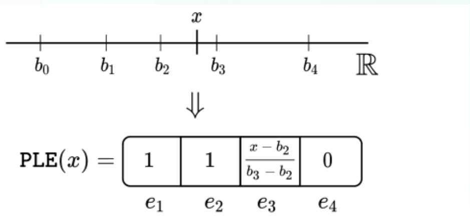
ET-RNN: Applying Deep Learning to Credit Loan Applications [KDD ’19]

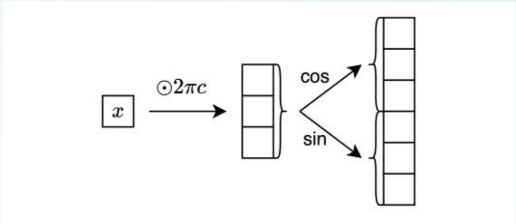
6

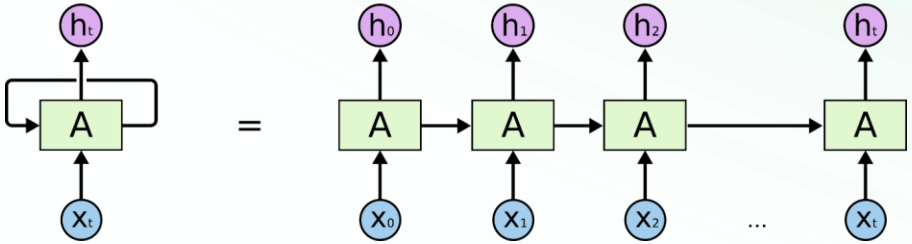
Architectures and Algorithms

Architectures:

- Transaction Encoders
- Sequence Encoders

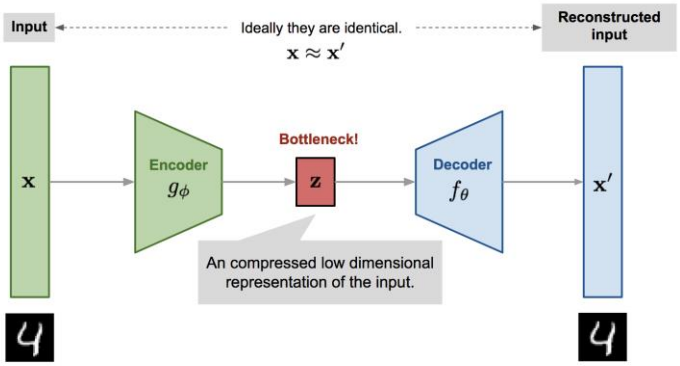




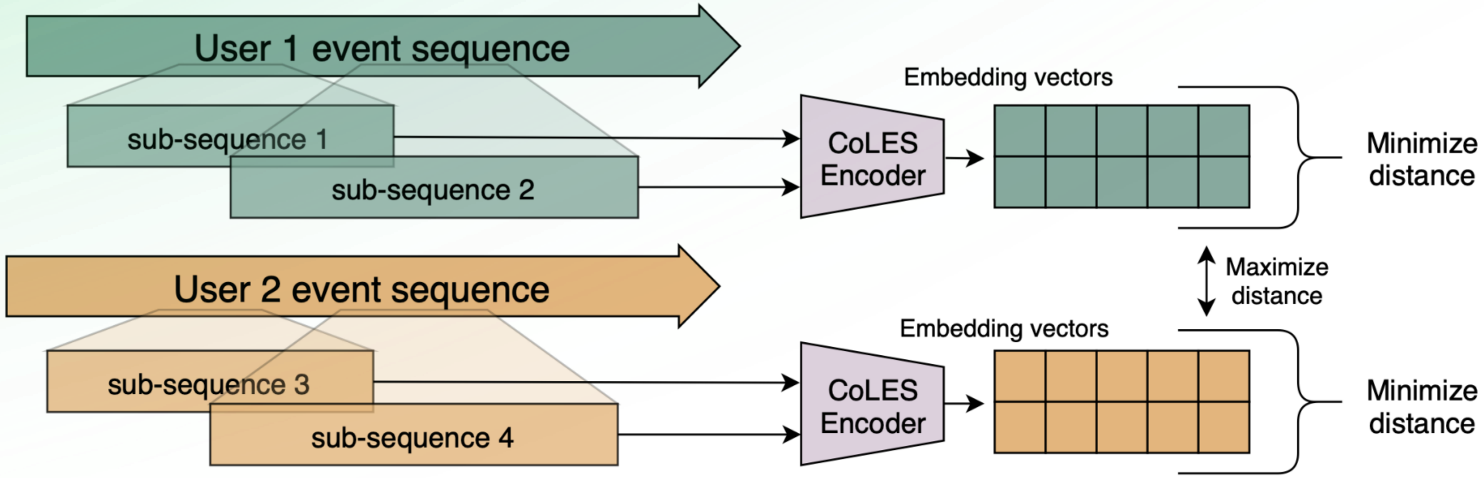


Learning Methods:

- Unsupervised / Self Supervised
- Contrastive



CoLES



CoLES: Contrastive Learning for Event Sequences with Self-Supervision
[SIGMOD'22]

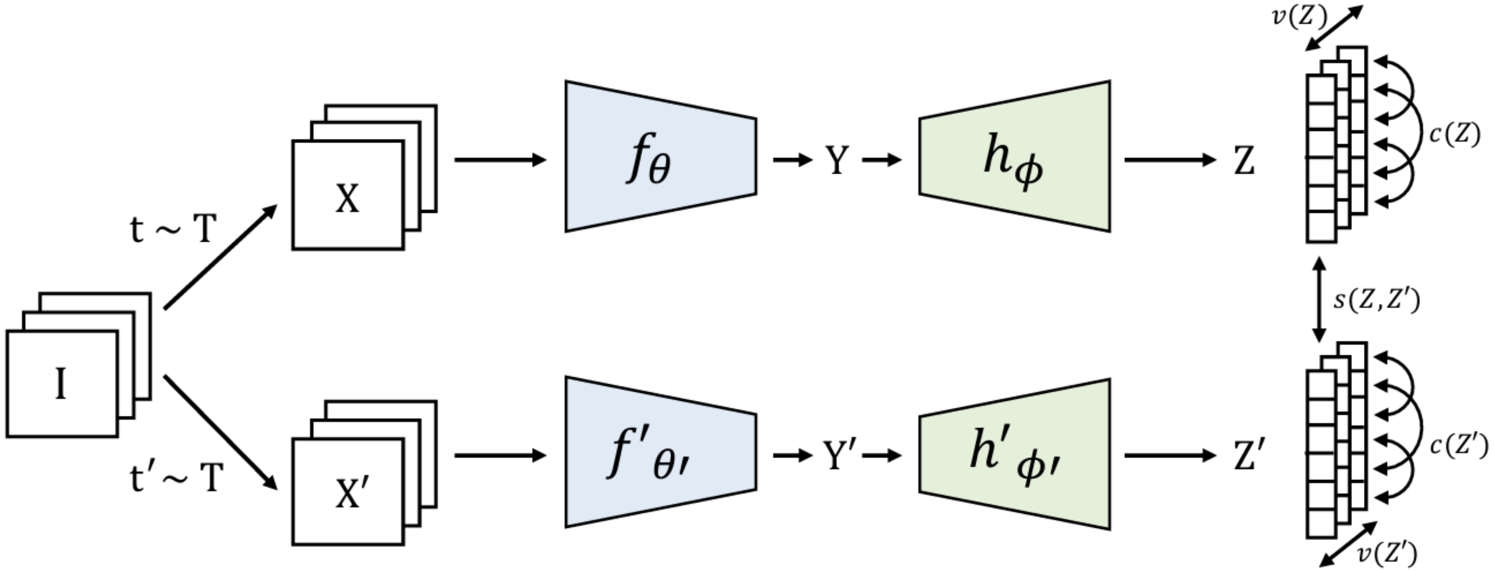
Advantages:

- Representation of the entire object
- Proximity is explicitly defined
- Splits resemble the entire sequence

Disadvantages:

- Negative examples are required
- Object dynamics are not taken into account

VICReg



VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning
https://arxiv.org/abs/2105.04906

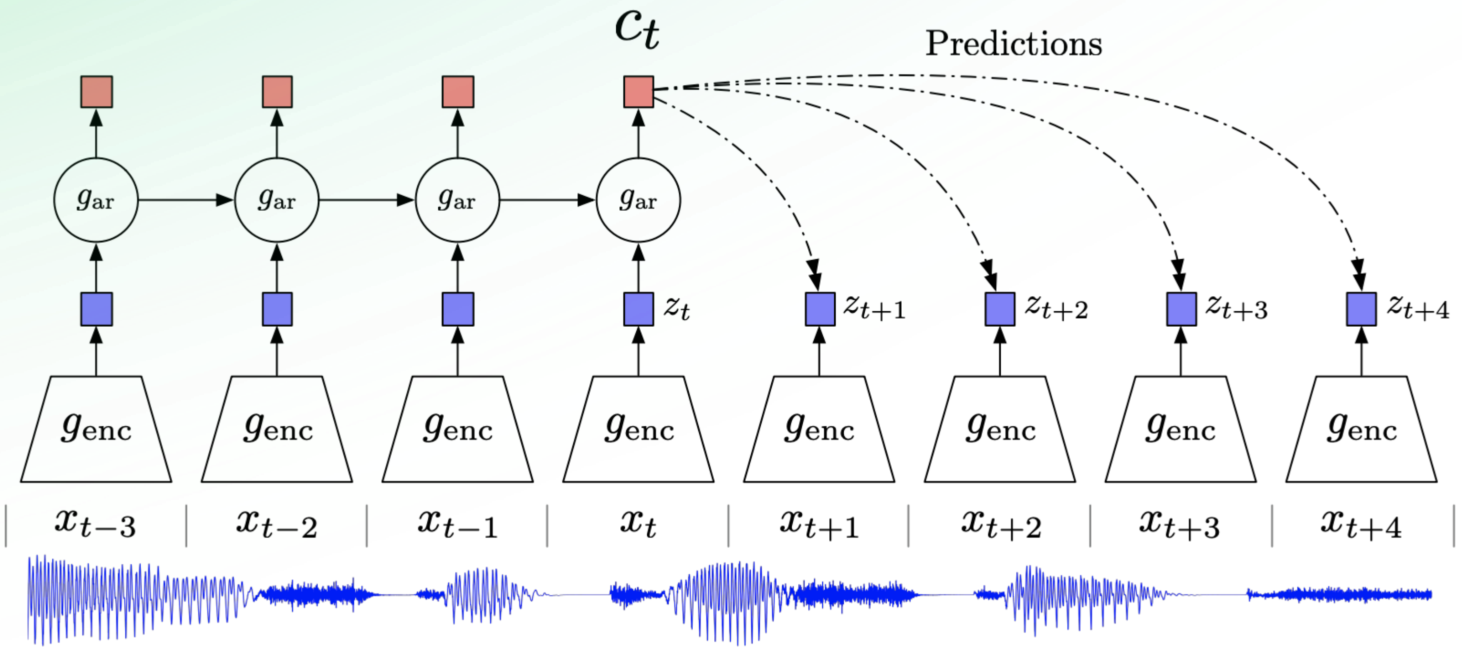
Advantages :

- Representation of the entire object
- Proximity is explicitly defined
- Negative examples are not required

Disadvantages :

- Sensitivity to hyperparameters

Contrastive Predictive Coding



Representation Learning with Contrastive Predictive Coding

https://arxiv.org/abs/1807.03748

Advantages :

- The hidden state contains all information about the object
- Predictive tasks are addressed

Disadvantages :

- Negative examples are required
- Mandatory splits are needed
- Predictive tasks are more complex

Multimodality for Embeddings

The use of additional data (modalities) enhances the quality of customer embeddings

Multimodal embeddings can be applied to the same tasks as traditional embeddings but perform better

Examples of Modalities

- Purchase History
- Financial operations
- Transfers
- Customer Communications
- Website and App Activity
- Receipts

Improvements for Individual Sources

- Raw, noisy data
- Large category dictionaries
- Rare events with limited coverage

New Types of Data

- Geolocation Data
- Graphs
- Text

Multimodal Banking Dataset

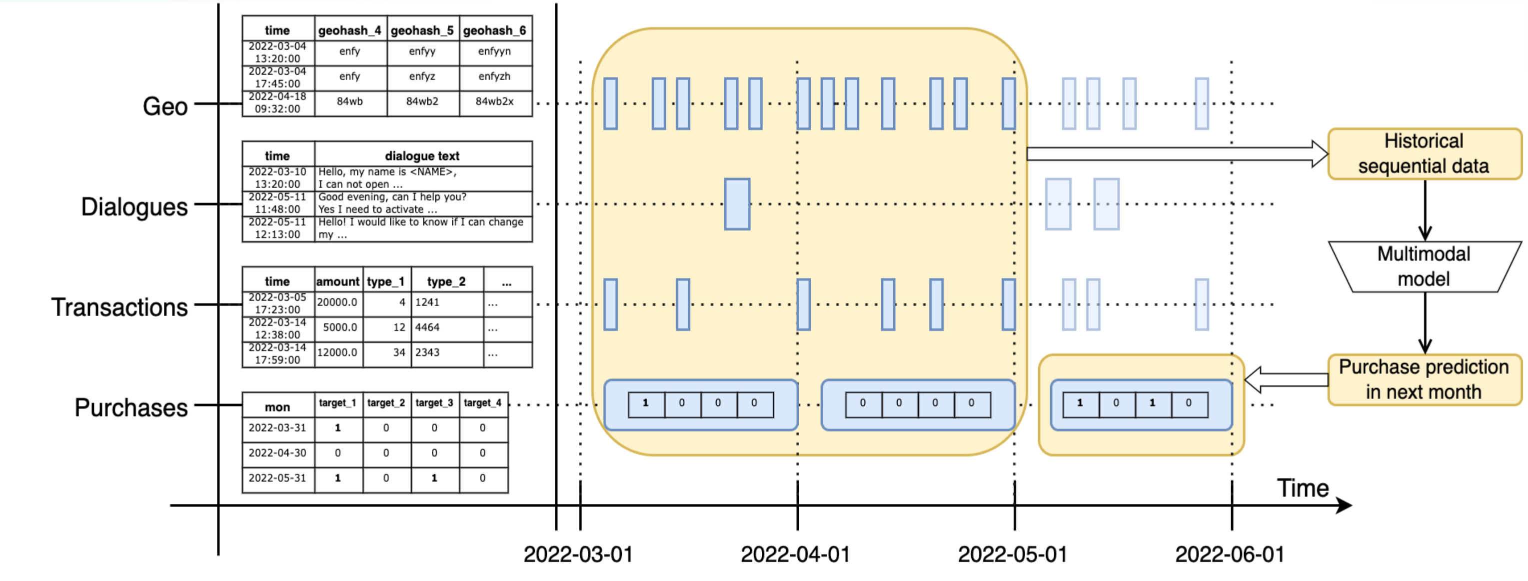
The largest open-source multimodal banking dataset

Data from 2 million clients has been collected and anonymized

Modalities:
- Transactions
- Dialogues
- Geostream

Task:
Predicting the purchase of 4 products for the next month

Link on Hugging Face:



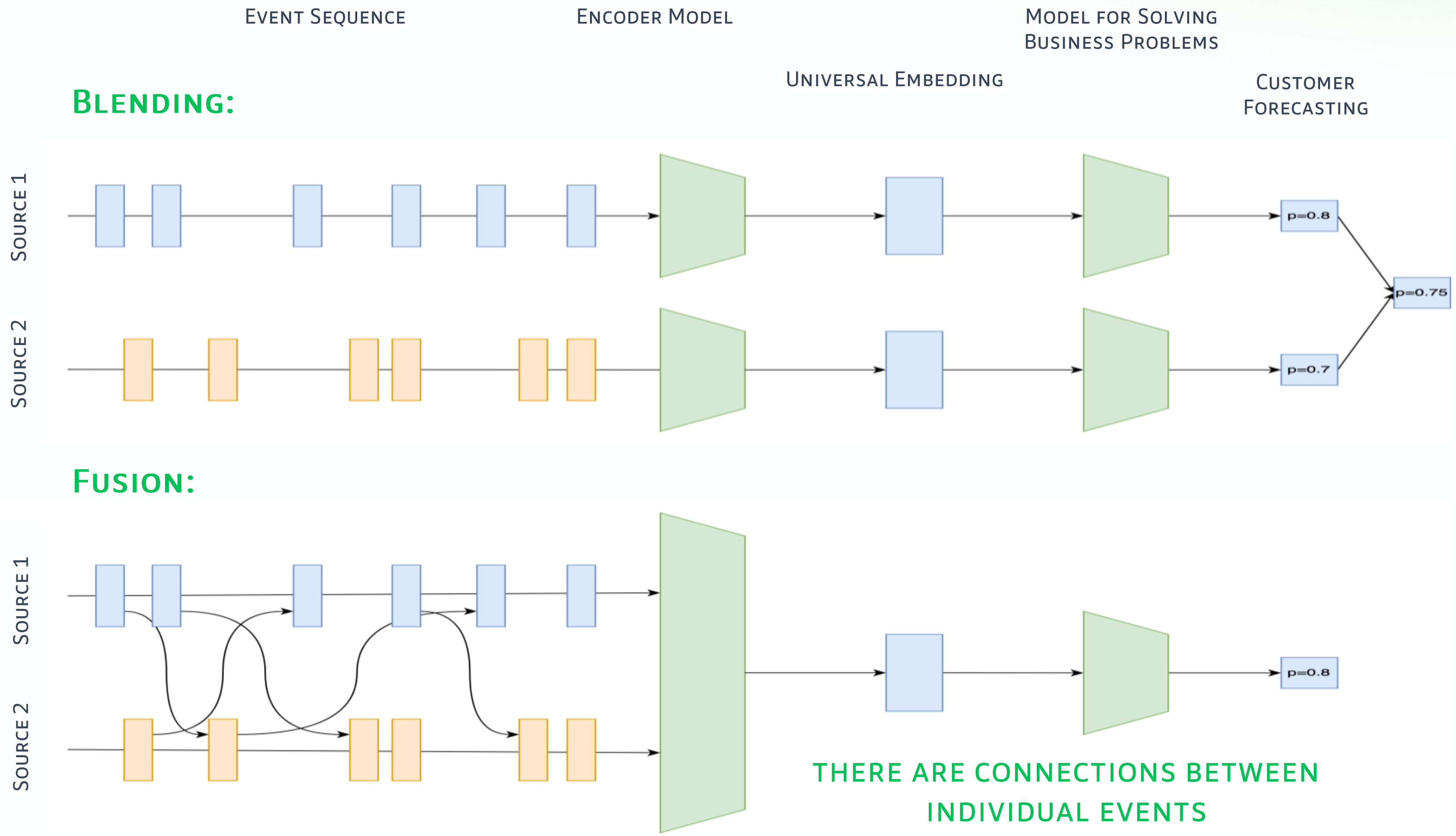
https://arxiv.org/abs/2409.17587

Deeper utilization of additional data results in higher quality

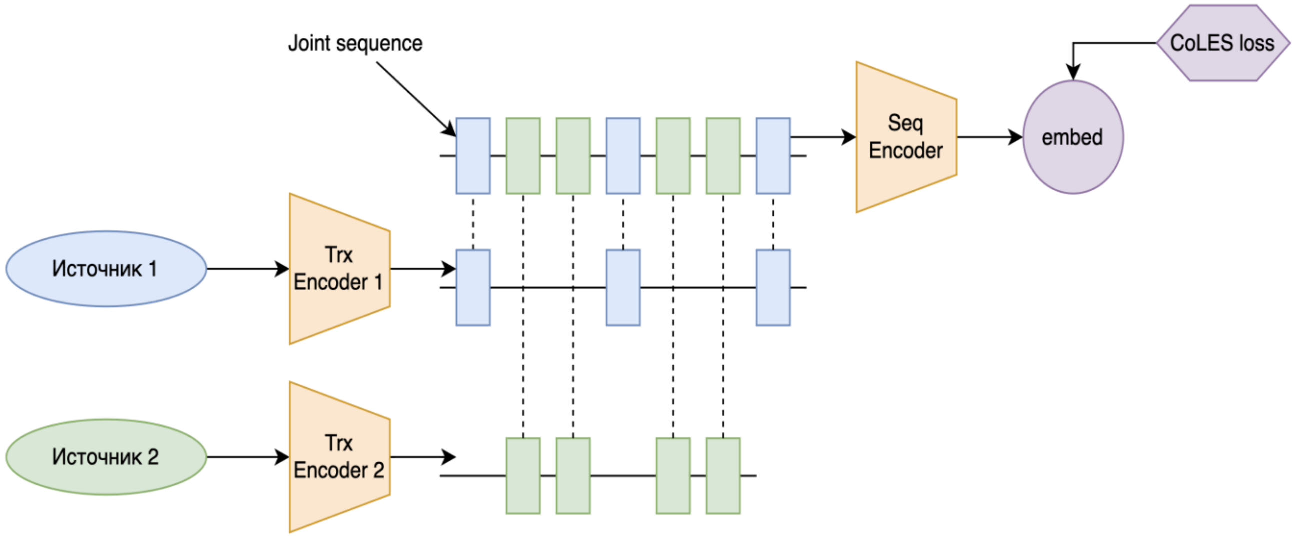
Options for Combining Modalities:

- Without using additional data
- Blending
- Late Fusion
- Early Fusion
- Mid Fusion

Fusion of Modalities - Accounting for Deep Relationships



Early Fusion
Объединение событий в одну цепочку



Description:

Events from each modality are mixed into a single chain

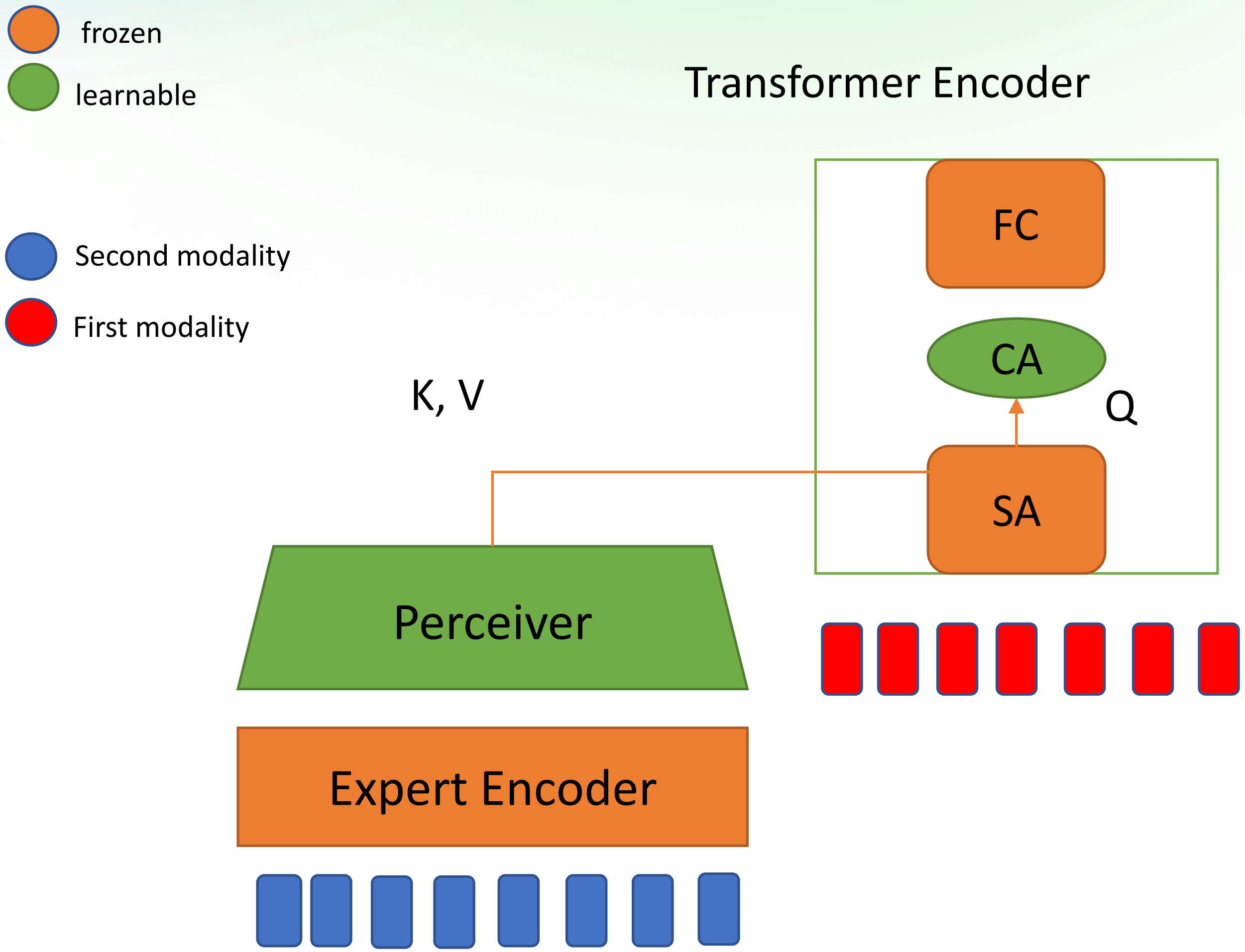
Advantages :

Allows for more detailed information about modalities than Late Fusion

Disadvantages:

Events with low frequency may get lost in the overall flow of events

Early Fusion Flamingo



Flamingo: a Visual Language Model for Few-Shot Learning, NeurIPS 2022

Description:

Incorporating modalities into the Transformer using cross-attention

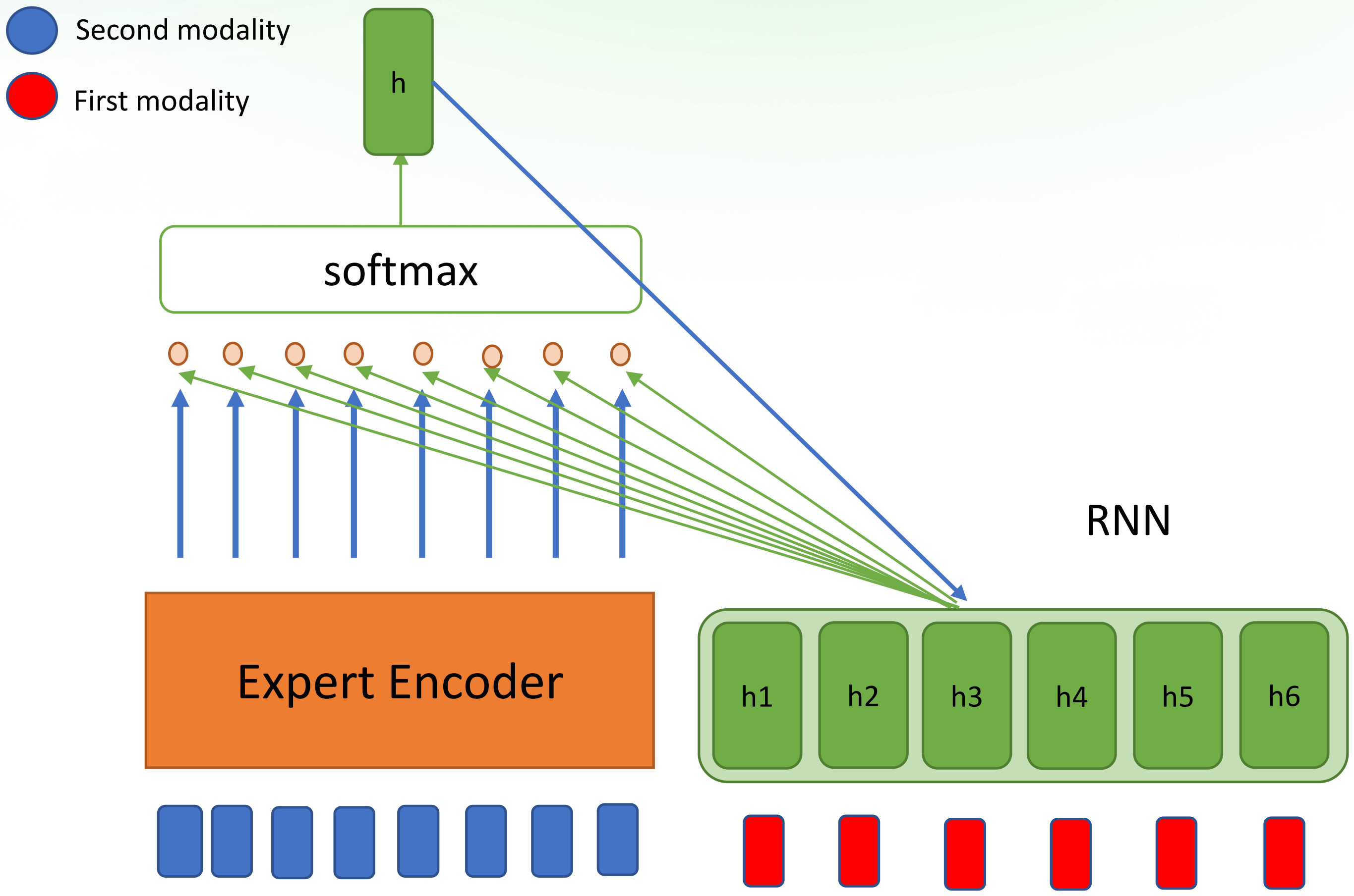
Advantages:

- Pre-training expert encoders for modalities allows for extracting more detailed information about individual modalities
- Trainable cross-attention in the Transformer layers helps address the issue of low-frequency modalities

Disadvantages :

- Scalability issues with a larger number of modalities
- Complex tuning of the Transformer for the event sequence domain.

Early Fusion Attention-Rnn



NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE, ICLR 2015

Description:

Embedding modalities in RNNs using attention

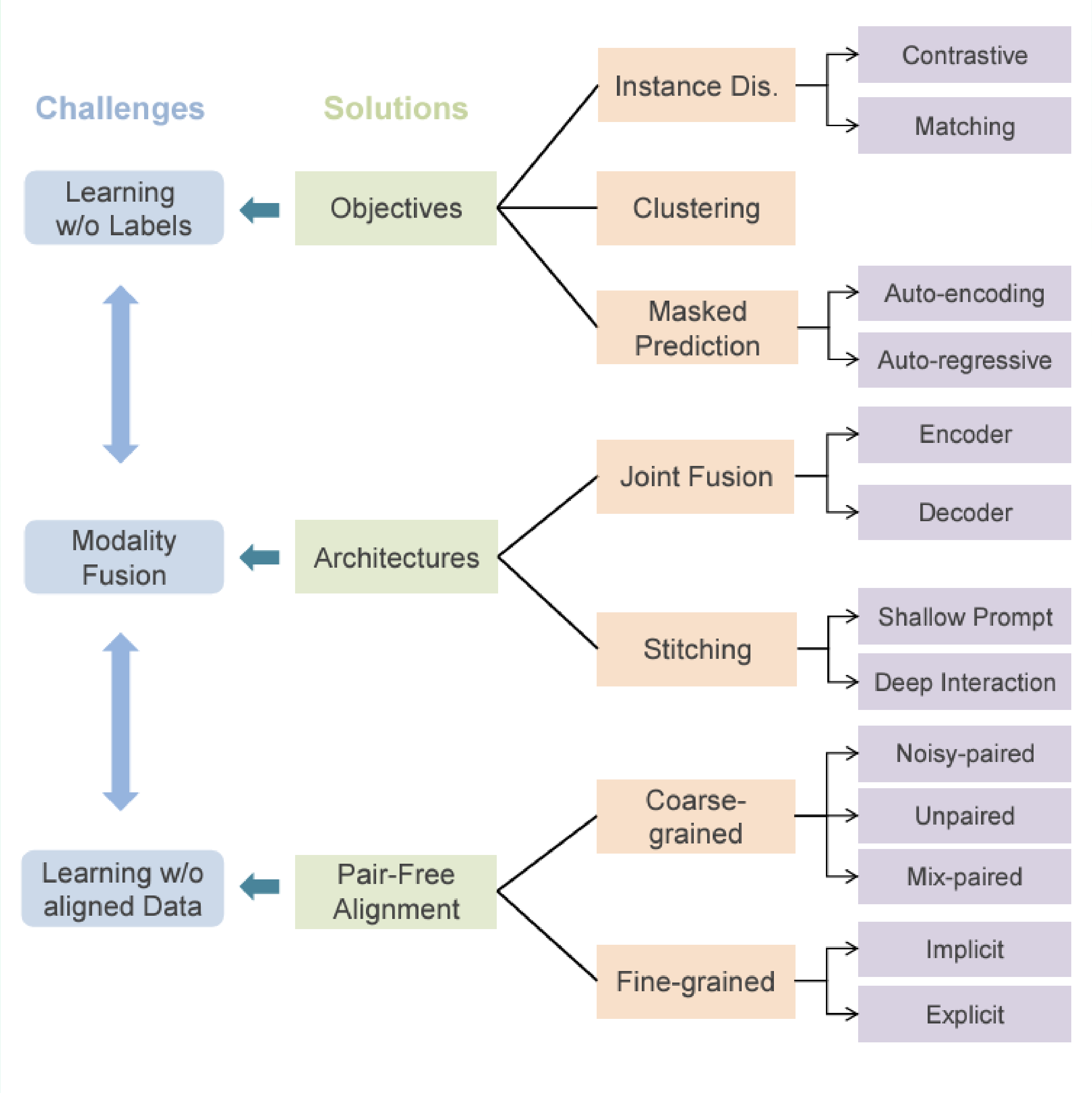
Advantages :

- Fewer parameters to tune
- RNN architecture performs better than Transformer for event sequence domains

Disadvantages :

Scalability issues with a larger number of modalities

Self-Supervised Multimodal Learning



Self-Supervised Multimodal Learning: A Survey

https://arxiv.org/pdf/2304.01008

Thank you for your attention!

GitHub
sb-ai-lab

Website
Sber AI Lab

@ivkir8

Ivan Kireev