WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

Training Foundation Models for Predicting Customer Behavior

Barbara Rychalska Dominika Basaj

SUMMIT.DATAMASS.IO

The problem of behavioral modeling

Product buys

event_uuid	client_id	event_date	event_timestamp	sku	name	shop	regular_price	label	quantity
8f69bb8b	730770875	2021-06-02	2021-06-02 15:41:14	10927568	BATON 3BIT 46G	x7632	2.60	NaN	1.0
fe8a460e	730770875	2021-06-02	2021-06-20 15:41:14	10901642	WAFEL KNOPPERS 25G	x7632	2.20	NaN	1.0
a5ecfa5a	916545677	2021-06-03	2021-06-03 09:26:39	12003716	SMOOTHIE FOODINI 250ML	x3187	4.99	NaN	1.0

Client properties

client_id	created_at	updated_at	geo_loc_status	geo_loc_timezone	tags
730770875	2021-11-02 11:51:14	2021-11-12 12:24:29	NaN	NaN	["test"]
916545677	2019-05-25 10:09:05	2021-11-12 12:23:34	NaN	NaN	["akcja_app_2020_08", "przekroczenie_limitow"
128788882	2019-05-25 10:09:05	2021-11-12 12:24:05	NaN	NaN	["test"]

The problem of behavioral modeling

The data is usually very varied...

Web interactions	Page views, searches, transactions, product returns, support queries, ...
Offline interactions	Transactions, contracts signed, customer support calls, ..
Mobile app interactions	Clicks, scrolls, push events, location, sensor data, ...
Financial	Invoicing, credit card payments, wire transfers, ATM withdrawals, ...
Telecom	Phone calls, text messages, internet usage, ...

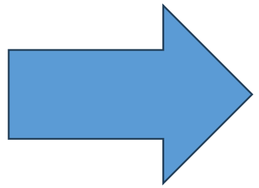
Customer attributes	Page views, searches, transactions, product returns, support queries, ...
Product attributes	Brands, titles, descriptions, colors, sizes, styles, images, …
employee attributes	seniority, specialty, skills, ….
marketing action attributes	Channel, format, description, text content, ….
financial	merchant categorization, merchant description, transfer destination metadata, ...

and many more

The problem of behavioral modeling

… and increasingly stored in the cloud

On-prem



snowflake

Google BigQuery

amazon REDSHIFT

…

Why?

- Security governance
- Scaling
- Efficiency (can store denormalized data)
- Availability of AutoMLs

Data Redundancy in the Cloud

In traditional on-prem databases, tables were often normalized to avoid redundancy.

In most cloud databases, storage is less expensive than compute

- In order to do ML, it is usually necessary to join the data
- However, joins are slow
- Fortunately, cloud storages allow for data to stay in the denormalized form and do ML much more quickly and easily

UserID	ProductID	Quantity	Timestamp
12780a4bc1234d	4567ef890ab2cd	2	2023-06-10 09:15:32
9e3f6d2a1b5c8f	2c4d6e8f0a1b3c	1	2023-06-11 17:42:19
8f9e6d5c4b3a2f	a1b2c3d4e5f6g7	3	2023-06-12 11:20:05
7g6f5e4d3c2b1a	123456789abdcef	1	2023-06-12 14:35:11

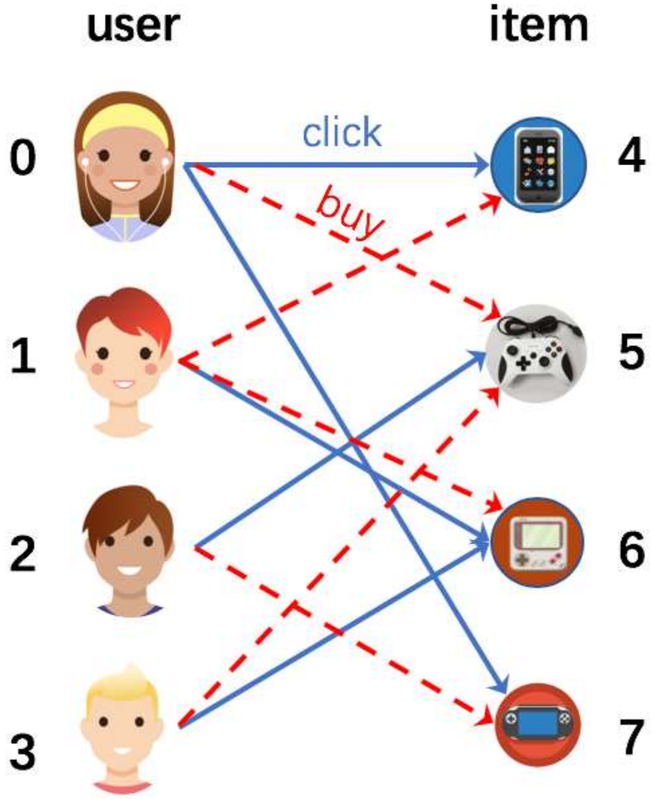
ProductID	ProductName	ProductCategory	ProductBrand
4567ef890ab2cd	Bananas	Fruits	Chiquita
2c4d6e8f0a1b3c	Milk	Dairy	OrganicFarms
a1b2c3d4e5f6g7	Bread	Bakery	WonderBread
123456789abdcef	Apples	Fruits	HoneyCrisp

UserID	ProductID	ProductName	ProductCategory	ProductBrand	Quantity	Timestamp
12780a4bc1234d	4567ef890ab2cd	Bananas	Fruits	Chiquita	2	2023-06-10 09:15:32
9e3f6d2a1b5c8f	2c4d6e8f0a1b3c	Milk	Dairy	OrganicFarms	1	2023-06-11 17:42:19
8f9e6d5c4b3a2f	a1b2c3d4e5f6g7	Bread	Bakery	WonderBread	3	2023-06-12 11:20:05
7g6f5e4d3c2b1a	123456789abdcef	Apples	Fruits	HoneyCrisp	1	2023-06-12 14:35:11

This would usually end up in a separate table to avoid redundancy. However, in the cloud it is usually left redundant.

Why current cloud AutoMLs are not enough

- **Behavioral Modeling is in fact Temporal Graph Processing**
  - Think GNNs, TGNs, Message Passing frameworks, Graph Embeddings
  - If we cannot use graph models, we must use aggregated semi-handcrafted features
- **But, existing AutoMLs do not explicitly support this type of modeling!**
  - They are focused on ordinary row datasets, or text/image modeling
- **Validation and testing must also be done in the temporal domain**
  - Option 1: test on user holdout sets
  - Option 2: test on the same users, but look at their future behaviors
- **User representations are hard to compute on-the-fly – usually they are precomputed (and are not fully up to date)**



What can we do?

Build platforms dedicated for graph learning, behavioral modeling and event modeling!

Snowflake Containers

- One click start – infra ready to go in Terraform
- Easy connector to BQ and other storage on cloud

Vertex Notebooks

- Easy to start – we can start out of the shelf in a dedicated container with BaseModel installed
- Easy to interact and explore because of our onboarding tutorials

Vertex Pipelines

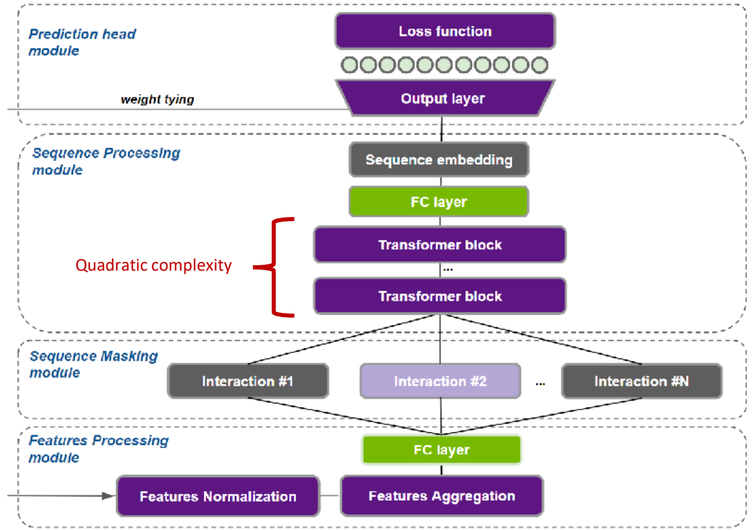
- Optimized, because Compute Engines are running only on request

Azure

- Model registry and easy deployment of ML models

So, is it enough if we implement a GNN, package it, and run in the cloud? Not really...

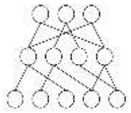
Transformer & GNN Philosophy
(on the example of Transformers4Rec)



Throw in data and let the network figure everything out

BaseModel Philosophy

Sparse feed-forward network

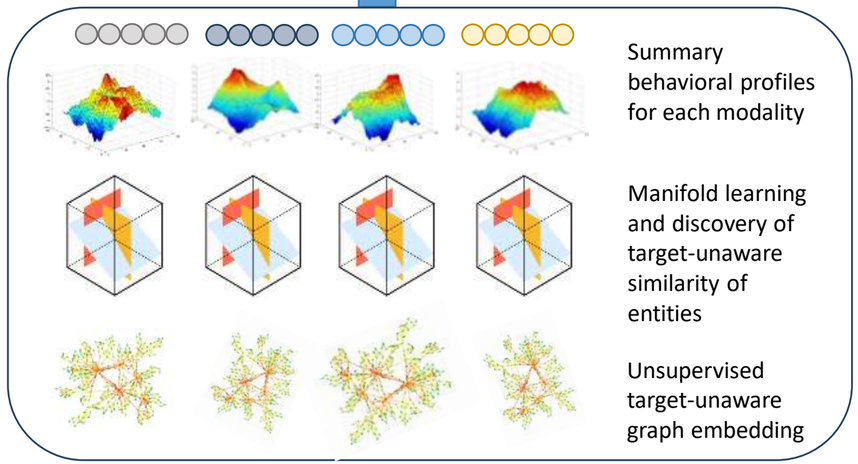


Modulate user summary vectors by trainable temporal and contextual information vectors

User Behavioral Representation

Linear complexity

Maximum amount of **structured** target-unaware information is fed into the network



Summary behavioral profiles for each modality

Manifold learning and discovery of target-unaware similarity of entities

Unsupervised target-unaware graph embedding

Create useful priors for the network

Unsupervised and target-unaware – good for all models and can be done only once

KDD2021
ACM KDD Cup Stanford OGB-LSC 2021

Booking.com
ACM WSDM Booking.com Data Challenge 2021

ACM RecSys Twitter Challenge 2021

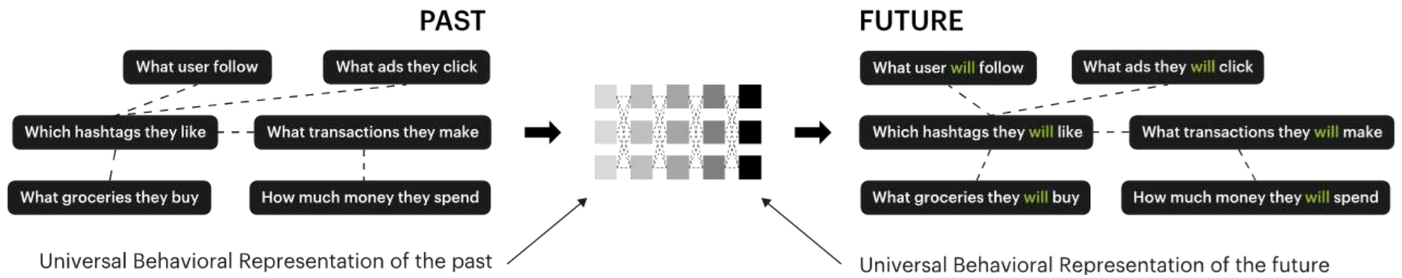
Rakuten
SIGIR eCom Rakuten Challenge 2020

How to additionally save on compute and modeling time

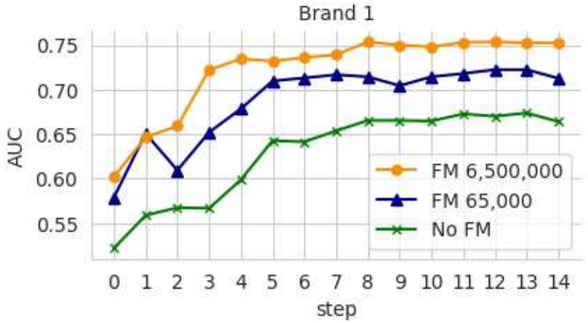
Use a foundation model.

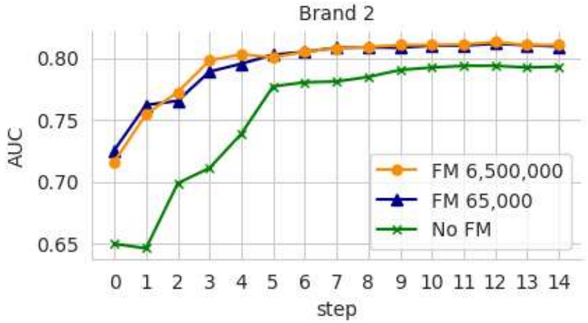
Foundation models are large, general-purpose models trained on big amounts of data

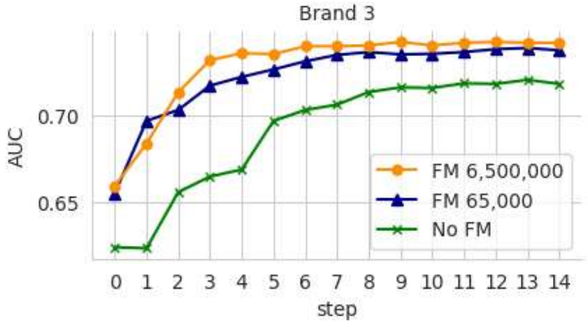
- Often equated with LLMs or generative models, but in fact Foundation Models (Base Models) can serve any purpose, and be also predicive
- Foundation Models are trained once and fine-tuned to specific tasks
- Fine-tuning is much faster than regular, full training



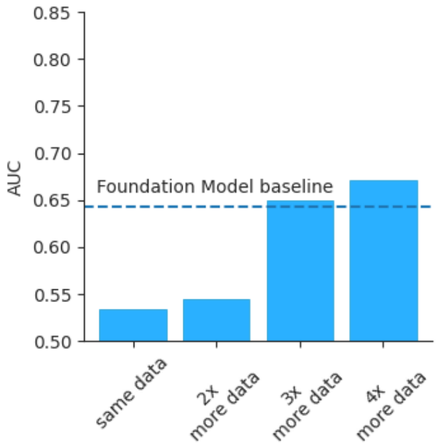
Behavioral Foundation Model

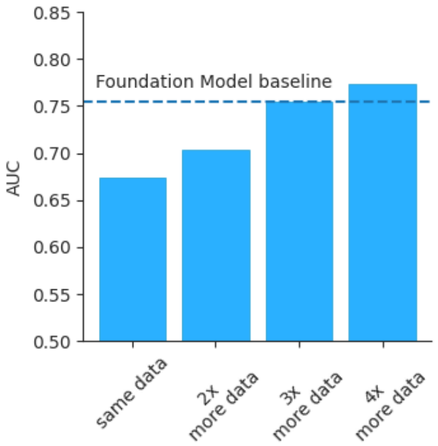






Prediction of propensity towards 3 brands – speed and data reqirements with and without a Foundation Model





Thank you

Barbara Rychalska

AI Research Director
Synerise/BaseModel.ai

Dominika Basaj

Head of Applied Data Science
Synerise/BaseModel.ai