MONAD is a private foundation model for behavioral data

Foundation models like ChatGPT, GPT-3, Dall-E 2, StableDiffusion have revolutionized Text and Image processing. A single large model trained on massive datasets can replace thousands of specialized models. For the first time, MONAD allows to apply the same principle to behavioral data.

© Synerise 2023
Reduce your modeling life-cycle to days instead of months

Understanding complex & intricate patterns of interactions is a super-human challenge. Imagine a single model could learn from all your raw data. Such a model could form a foundation for solving any applied task with unparalleled efficiency and quality. **This is exactly what MONAD does.**

Until now:
- each ML project required careful manual labor, starting with analyzing available data sources
- countless handcrafted features had to be created using expert knowledge
- despite best efforts, important behavioral cues were often lost due to human limitations
- the information content of raw data was orders of magnitude richer than the actual input of models

MONAD eliminates these problems and supercharges behavioral ML.

**Current standard approach**

<table>
<thead>
<tr>
<th>Understand data</th>
<th>Identify entities, attributes</th>
<th>Find good joins</th>
<th>Filter data</th>
<th>Clean data</th>
<th>Creature features</th>
<th>Identify useful features</th>
<th>Automate feature transforms</th>
<th>Create target variables</th>
<th>Train &amp; tune models</th>
<th>Add more features</th>
<th>Retrain models</th>
<th>Discuss results with business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Monad**

<table>
<thead>
<tr>
<th>Connect to data</th>
<th>Fit foundation model</th>
<th>Train models</th>
<th>Discuss results with business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Just 5 days: from MONAD configuration to first trained model**

Deploy Monad
- Cloud or on-premise deployment
- Docker Container
- 1 VM with 256 GB RAM
- 1 Nvidia GPU

Retrieve data

Foundation model training
- Self-supervised behavioral foundation model

Fine tuning
- Fine-tuned supervised models for specific use-cases

Synerise Monad - Apply science to behavioral data. Automatically.
MONAD: unprecedented accuracy of behavioral modeling

MONAD looks at all human interactions with your organization and learns to predict future behaviors. Automatically. At scale.

Example:

For MONAD, solving applied problems becomes a piece of cake thanks to understanding the full spectrum of behaviors. The foundation model is trained only once, and then adapted to specific tasks in a process called “fine-tuning”. Not having to re-learn behaviors from scratch, a fine-tuned model only needs a target to zoom in on.

Your private foundation model can bootstrap business applications like:

- propensity prediction
- churn prediction
- recommendations
- anomaly detection
- customer scoring
- customer matching
- behavioral hyper-segmentation
- and many more

Just like large language models can be fine-tuned to any application, **MONAD is not restricted to a predefined set of use cases** – you can unleash your creativity freely.

No more „plumbing pipelines” and „handcrafting model inputs”. Ideate, prototype, evaluate!

Thanks to MONAD, applied models train faster, require less labelled data and perform better than classic approaches.

- MONAD models are trained **2x faster** than baselines
- MONAD needs **3.5x less** labelled data than baselines
- MONAD’s quality metrics **significantly** outperform classic baselines
- MONAD performs advanced feature engineering, representation learning & training **automatically**
Challenging top AI powerhouses

We take science seriously. We verified MONAD’s power at the most prestigious research venues, where early versions of MONAD took the podium competing with global giants like Google DeepMind, Baidu Research, NVIDIA, Intel, Oppo Research, Xerox PARC, Rakuten and 800+ teams from leading universities worldwide.

The difference? Our competitors spent weeks hand-crafting their solutions for every problem separately, with zero re-usability. We just applied MONAD to all the problems.

Example questions MONAD can answer

<table>
<thead>
<tr>
<th>Organizer</th>
<th>Competition name</th>
<th>Description of the challenge</th>
<th>1st Prize</th>
<th>2nd Prize</th>
<th>3rd Prize</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KDD2021</strong></td>
<td>ACM KDD Cup Stanford OGB-LSC 2021</td>
<td>Predict the subject of scientific publications on the basis of edges contained in the heterogeneous graph of papers, citations, authors and scientific institutions.</td>
<td>Bar</td>
<td>Google DeepMind</td>
<td>Synerise Monad</td>
</tr>
<tr>
<td><strong>Booking.com</strong></td>
<td>ACM WSDM Booking.com Data Challenge 2021</td>
<td>Provide the best suggestion for the next trip most likely to be bought by a booking.com customer based on millions of real, anonymized bookings of existing clients.</td>
<td>NVIDIA</td>
<td>Synerise Monad</td>
<td>Team Dasou</td>
</tr>
<tr>
<td></td>
<td>ACM RecSys Twitter Challenge 2021</td>
<td>Based on 1 billion historical tweets predict 4 possible actions (likes, retweet, comment, retweet with comment) of the next tweet provided in a random language.</td>
<td>NVIDIA</td>
<td>Synerise Monad</td>
<td>layer 6</td>
</tr>
<tr>
<td><strong>Rakuten</strong></td>
<td>SIGIR eCom Rakuten Challenge 2020</td>
<td>Identify the images of a huge data base of products without being given pairing of descriptions and images.</td>
<td>Synerise Monad</td>
<td>Undisclosed</td>
<td>Rakuten Institute of Technology</td>
</tr>
</tbody>
</table>
Technical Appendix

Examples of behavioral data MONAD can be fueled with

Events:
- **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, ...

Attributes and metadata:
- **customer attributes**: socio-demographic, location, loyalty program, subscriptions, ...
- **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, ...

Thanks to extremely efficient algorithms and optimizations, MONAD is blazingly fast. No clusters needed

Our mission to equip every organization with a private behavioral foundation model makes us very conscious about resources. That’s why we take cost-efficiency seriously & optimize our algorithms to be lightning-fast and extremely scalable.

A full self-supervised training pipeline for 10M+ customer profiles with 1 year worth of data can complete within a few hours on a VM equipped with a single GPU. Subsequent fine-tuning for supervised applications is even faster, and requires very little labeled data.

<table>
<thead>
<tr>
<th>Datasource</th>
<th>Timespan</th>
<th>Number of rows</th>
<th>Fitting time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events 1</td>
<td>1 year</td>
<td>&gt; 3 billion</td>
<td>100 minutes</td>
</tr>
<tr>
<td>Events 2</td>
<td>1 year</td>
<td>~ 630 million</td>
<td>50 minutes</td>
</tr>
<tr>
<td>Client Attributes</td>
<td>-</td>
<td>10 million</td>
<td>20 seconds</td>
</tr>
</tbody>
</table>

Simple, secure and quick deployment, with zero overhead

We’ve made sure that starting the journey with MONAD is smooth:

- MONAD can be deployed on your own **cloud subscription**, or on-premise (NVIDIA GPU-equipped server required)
- MONAD is available as a hardened **Docker image**. It doesn’t require clusters, multiple micro-services or standalone databases.
- MONAD does not duplicate your data. It reads the necessary data **directly from your data-warehouse** in a streaming way, during model training & inference.
- Thanks to the above requirements, MONAD is easily auditable and can pass the most **stringent security requirements**.
How does MONAD work under the hood?

MONAD automatically finds proper representations suitable for aggregation of data, such as:

- graphs
- texts
- images
- numbers
- categorical variables

It utilizes a mix of Graph ML, differential geometry and Deep Learning.

MONAD uses proprietary research to represent complex multi-modal, multi-source histories of behavior in the form of sparse vectors, called Universal Behavioral Representations. Technically, these vectors represent probability density estimates over Riemannian product manifolds and can serve as both inputs and targets for neural network training. In simpler words, MONAD compresses multi-modal event series into very wide fixed-length sparse vectors.

The key property of MONAD’s representations is that they are approximately reversible - which means, that it is mathematically possible to query a Universal Behavioral Representation about the elements aggregated within, with high accuracy. This allows neural models to “ask” very specific questions about the user’s historic activities, without the need for encoding precise knowledge as manual features. This unique property also allows for fine-grained interpretability of models (down to the lowest level of raw data).

Cleora + EMDE

Some of the algorithms used in MONAD like Cleora and EMDE have been open-sourced and can be found in the last section of the Appendix.

We’re committed to scientific transparency by publishing and open-sourcing parts of our research

OSS Cleora (graph + hypergraph embeddings):

- **Cleora**
  - Scalable learning of stable and inductive entity embeddings for heterogeneous relational data
  - [http://cleora.ai](http://cleora.ai)

OSS EMDE (multi-modal behavioral fusion):

- **EMDE**
  - Efficient Manifold Density Estimator for high-dimensional vector spaces
  - [https://github.com/Synerise/kdd-cup-2021](https://github.com/Synerise/kdd-cup-2021)
  - [https://github.com/Synerise/booking-challenge](https://github.com/Synerise/booking-challenge)

Publications

- **Theory**
  - Theoretical underpinnings of technology behind MONAD
    - An efficient manifold density estimator for all recommendation systems
    - T-EMDE: Sketching-based global similarity for cross-modal retrieval
• I know why you like this movie: Interpretable Efficient Multimodal Recommender  

• Cleora: A Simple, Strong and Scalable Graph Embedding Scheme  
https://arxiv.org/abs/2102.02302

• On the Unreasonable Effectiveness of Centroids in Image Retrieval  
https://arxiv.org/abs/2104.13643

• A Strong Baseline for Fashion Retrieval with Person Re-Identification Models  

Application

MONAD applied to real life use cases

• Synerise at KDD CUP 2021: Node classification in massive heterogenous graphs  
https://ogb.stanford.edu/paper/kddcup2021/mag240m_SyneriseAI.pdf

• Synerise at RecSys 2021: Twitter user engagement prediction with a fast neural model  
https://arxiv.org/abs/2109.12985

• Modeling Multi-Destination Trips with Sketch-Based Model  

• Synerise at SIGIR Rakuten Data Challenge 2020: Efficient Manifold Density Estimator for Cross-Modal Retrieval  

• Synerise at SIGIR Rakuten Data Challenge 2020: Efficient Manifold Density Estimator for Multimodal Classification  