Knowledge Enhanced Deep Learning: Application to Pandemic Prediction

John A. Miller
Nasid Habib Barna
Subas Rana
I. Budak Arpinar
Ninghao Liu

University of Georgia
School of Computing
Athens, GA
USA

Outline

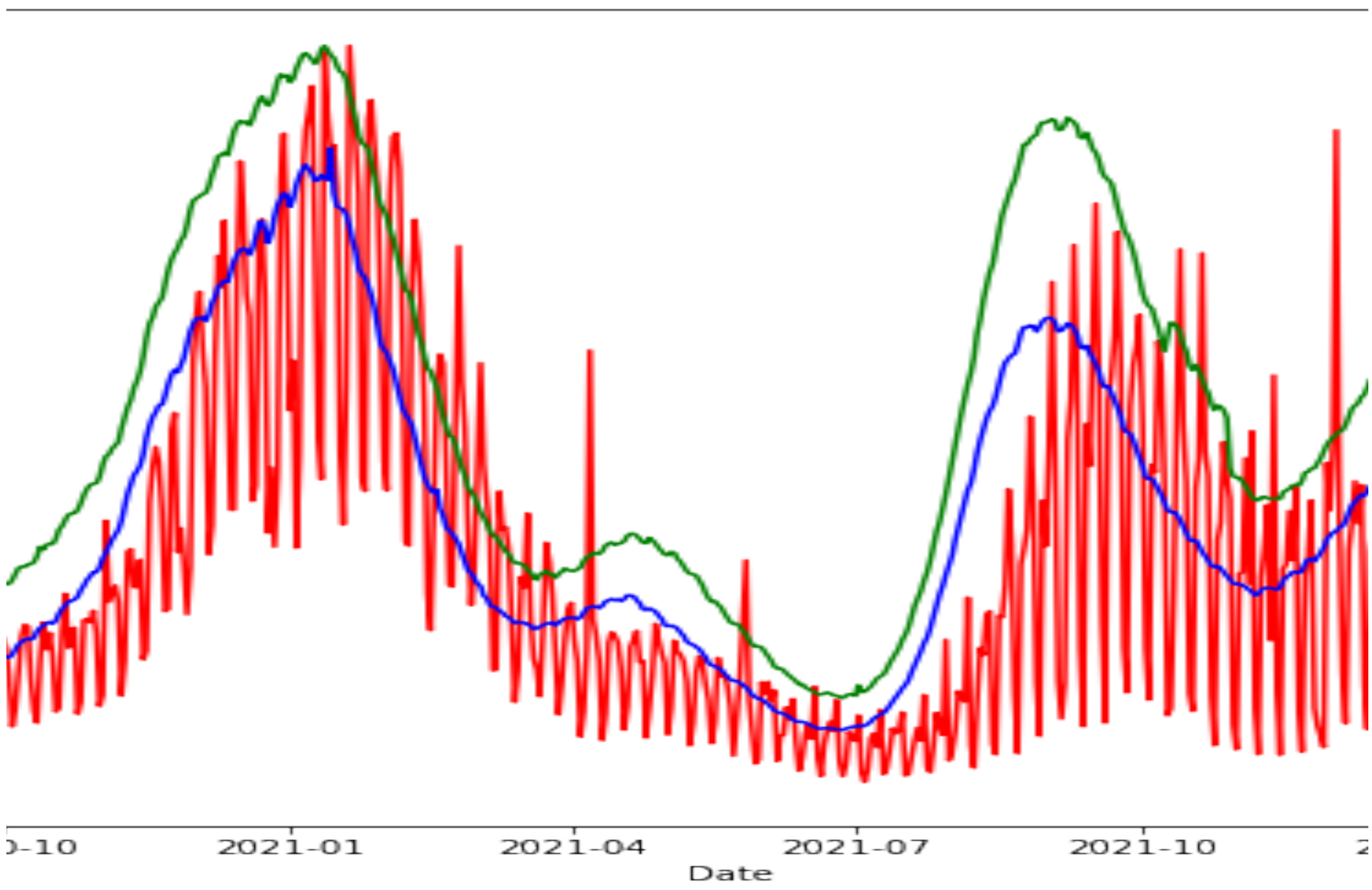
- Introduction
- Progress in Multivariate Time Series Forecasting
- Recent Progress on Transformers for Time Series
- Recent Progress on Graph Neural Networks for Time Series
- Use of Knowledge

Introduction

- Pandemic Prediction
- Forecasting from Emerging Patterns
  - Not enough data
- Using Similar Past Epidemics/Pandemics as Guides
- Applying Scientific Knowledge
- Combining Data-Driven and Theory-Driven Approaches

Rescaled Plot of Daily Deaths

deaths, hospitalizations, ICU



Progress in Multivariate Time Series

- Progress with Statistical Models
  - SARIMAX - ideal baseline for model comparison
  - Combinations & Ensembles (M5 competition winners)
- Progress with Machine Learning
  - Classic Machine Learning - LightGBM popular in M5
  - Deep Learning - RNN (GRU/LSTM), Encoder-Decoder, Transfomers, GNN

Recent Progress on Transformers

Full, Quadractic Attention; followed by more efficient and focused attention

1. **Vanilla Transformer** - Positional Encoding and Self-Attention
2. **LogTrans** - Local and LogSparse Attention
3. **Reformer** - Only Similar Queries and Keys Are Compared
4. **Informer** - Uses Selected Query Prototypes
5. **Autoformer** - Replaces Self-Attention with Auto-Correlation
6. **Pyraformer** - Hierarchical/Pyramidal Attention
7. **FEDformer** - Series Decomposition and Use of Frequency Domain
8. **Crossformer** - Exploits Inter-Variable (Cross) Dependencies
9. **AR-Transformer** - Segmented, Auto-Regressive Transformer

[Aldosari2023transformer]

Comparison of Transformers

MAE for Influenza-Like Illness (ILI Weeks)

Model	24	36	48	60	Rank
Vanilla	1.323	1.360	1.463	1.553	
LogTrans	1.444	1.467	1.468	1.560	
Reformer	1.382	1.448	1.465	1.483	
Informer	1.462	1.496	1.516	1.576	
Autoformer	1.238	1.270	1.203	1.202	3
Pryaformer	1.338	1.410	1.503	1.588	
FEDformer	1.147	1.160	1.155	1.163	2
Crossformer	1.186	1.232	1.221	1.305	3
AR-Transformer	0.998	0.872	0.924	0.992	1

Mean Absolute Error (MAE) of precent of patients with ILI

Recent Progress on Graph Neural Networks

1. **DCRNN** - Diffusion Convolution Recurrent Neural Network integrates graph convolution into an encoder-decoder gated recurrent unit

2. **STGCN** – Spatial-temporal graph convolutional network incorporates graph convolutions with 1D convolutions

3. **Graph WaveNet** - GraphWaveNet adds a self-adaptive adjacency matrix into graph convolution and employs wavenet frameworks

4. **GMAN** - A graph multi-attention network with spatial and temporal attentions

5. **MRA-BGCN** - A multi-range attentive bicomponent GCN that uses bicomponent graph convolutions with a multi-range attention mechanism

6. **MTGNN** – Multivariate time-series based GNN proposes a novel graph structure learning approach

7. **SLCNN** - Structure Learning Convolutional Neural Network defines structure learning convolutions and Pseudo three Dimensional convolutions to capture temporal dependence

8. **STFGNN** - Spatial-Temporal Fusion Graph Neural Network integrates a fusion operation with a gated convolutional module

Comparison of GNNS
MAE for Traffic Prediction (60 mins)

Model	METR-LA Dataset	PeMS-BAY Dataset	Rank
DCRNN	3.60	2.07	
STGCN	4.59	2.49	
Graph WaveNet	3.53	1.95	
GMAN	3.40	1.86	2
MRA-BGCN	3.49	1.91	3
MTGNN	3.49	1.94	
SLCNN	3.30	2.03	2
STFGNN	3.18	1.66	1

Traffic Speed

Use of Knowledge

- Knowledge vs. Data
- Forms of Knowledge
- Incorporation of Knowledge
- Knowledge Enhanced Deep Learning Models
    - Hope: improved accuracy, especially when training data are lacking and for longer term forecasting

Data vs. Knowledge

Conceptual hierarchy of knowledge

The generally accepted view sees [Tuomi1999data]

- **Data** as simple **facts** that become
  - e.g., value produced by a **sensor**
- **Information** as data is combined into meaningful **structures**, which subsequently become
  - e.g., put into a relational or graph **database**
- **Knowledge** as meaningful information is put into a **context** and when it can be used to make **predictions**
  - e.g., **recurring patterns**

Forms of Knowledge

- **Knowledge**
  - Recurring: likely to be exhibited in another place and time
  - Patterns: something discernible
- **Forms of Knowledge** [Harmelen2008handbook] - information enhanced/generalized
  - Knowledge Graph
  - Ontology based on some Description Logic
  - Rule-Based: Decidable subset of FOL, e.g., Horn Logic; Probabilitic Soft Logic
  - Equations: Algebraic, Difference, Differential, etc.
  - Natural Language
- Data in information structures are well **used by predictive models**, what about knowledge

Incoporation of Knowledge

1. Composite Loss Function

- $(1-\lambda)\|\mathbf{y}-\mathbf{y}\hat{}\|_p + \lambda\|\mathbf{z}-\mathbf{y}\hat{}\|_p$ where $\mathbf{y}$ = time series, $\mathbf{y}\hat{}$ = predicted values, $\mathbf{z}$ = predictions from theory-based model

2. Applying Constraints

- $\|\mathbf{y} - \mathbf{y}\hat{}\|_p + \lambda f_c(\mathbf{y}\hat{})$ where f_c is a penalty/regularization function based on constraint violation

3. Factored into Self-Attention Mechanism

- relevance of $\mathbf{y}_{tj}$ to $\mathbf{y}_{t-l,k}$ could be maintained in a temporal knowledge graph (variable j to k with lag l)
- used to focus or modify self-attention calculations.

Incoporation of Knowledge (2)

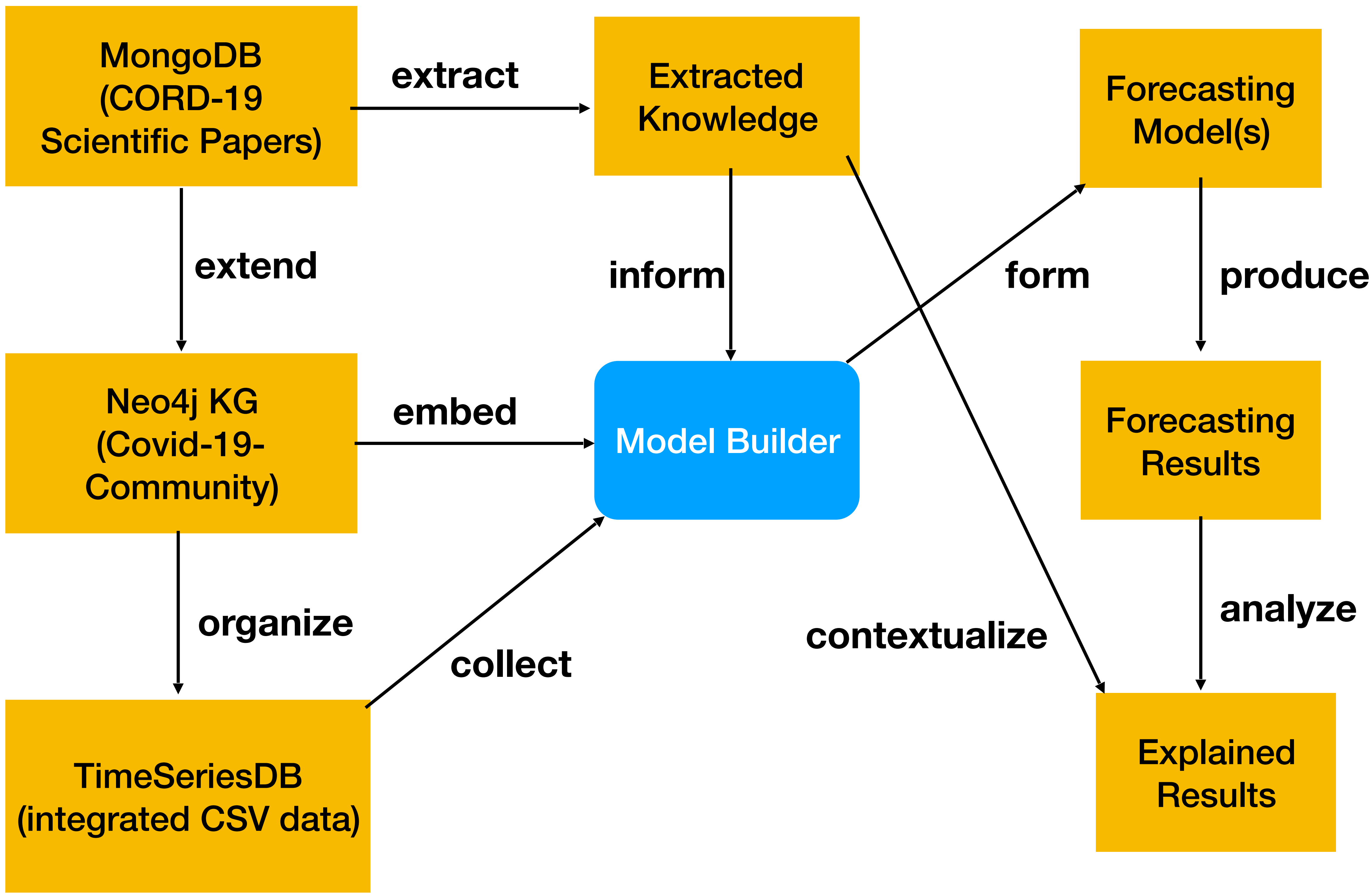
1. **Embedded and Combined with Input**
   - A sub-graph of a COVID-19 (Temporal) Knowledge Graph would produce embedded vectors
   - Combined/concatenated with input multivariate time series (e.g., raw/patch level).
2. **Injected into a Downstream Layer**
   - Determining ideal place to combine knowledge with input data or latent representations thereof is challenging.
   - Using representation learning that map $\mathbf{x}_t$ to $\mathbf{z}_t$, it could happen anywhere before the final representation is created.
3. **Knowledge Infuencing Architecture**
   - A sub-graph of a COVID-19 (Temporal) Knowledge Graph could also be used as a draft architecture for GNN.

Knowledge Enhanced Deep Learning Models

- **Data-driven methods** have made great strides of late
- But may still benefit by using **accumulated knowledge**
- How to use **scientific knowledge**
  - Theory - Epidemiology, Pharmacology, Genetics
  - Models - branching process model, compartmental models (SIR, SEIR) [Brauer2017mathematical]; KDS used in Knowledge Enhanced Neural Network (KENN)

Knowledge Enhanced Models

- (Temporal) **Knowledge Graph Embedding** (e.g., TTransE)
  - Concatenate with input $\mathbf{x}_{t;}$
  - Inject into a downstream layer
- **Neural ODE**
  - Use PyTorch's Automatic Differentiation with an ODE Solver to optimize an SEIR model's parameters
  - Encode the input, use ODE solver, decode its output
- **LLM Prompt** Engineering and Learning [Pryzant2023automatic]



Knowledge Enhanced Forecasting System

Summary

- Progress in **Time Series Forecasting** has been substantial
- **Deep Learning Models** can capture dependencies in time and space
- **Utilization** and **Integration** with **Scientific Knowledge** is being explored to make further progess.

Additional References

- Aldosari, M. and Miller, J. A. (Oct 2023). On Transformer Autoregressive Decoding for Multivariate Time Series Forecasting. In *Proceedings of the European Symposium on Artificial Neural Networks. ESANN 2023*.
- Tuomi, I. (Jan 1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32.* (pp. 12-pp). IEEE.
- Van Harmelen, F., Lifschitz, V., and Porter, B. (Eds.). (2008). *Handbook of knowledge representation*. Elsevier.
- Brauer, F. (2017). Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, *2*(2), 113-127.
- Pryzant, R. et al. (2023). Automatic Prompt Optimization with “Gradient Descent” and Beam Search, arxiv.