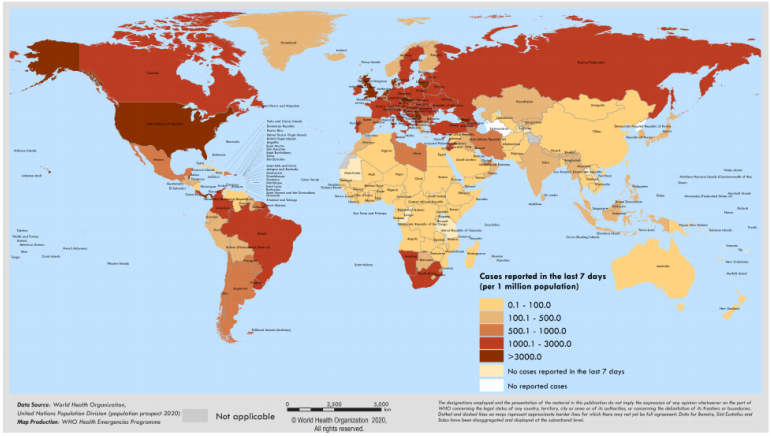
STELAR: Spatio-temporal Tensor Factorization with Latent Epidemiological Regularization

Nikos Kargas, Cheng Qian, Nicholas D. Sidiropoulos, Cao Xiao, Lucas M. Glass, Jimeng Sun

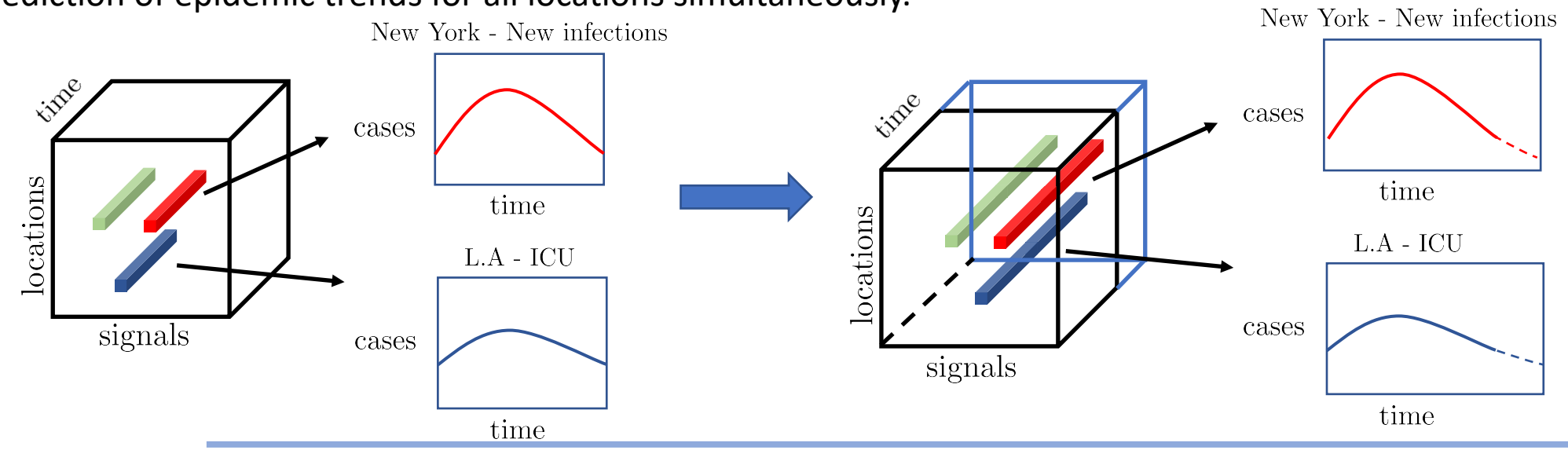
Epidemic Prediction

- Pandemic diseases
  - Serious threat to public heath, economy and daily life.
  - Accurate measurement, modeling and tracking are needed.
  - Effective mitigation measures.



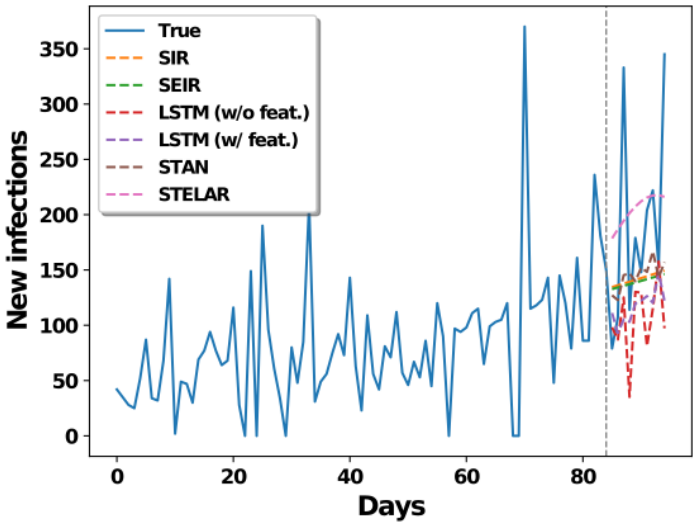
who.int December 2020

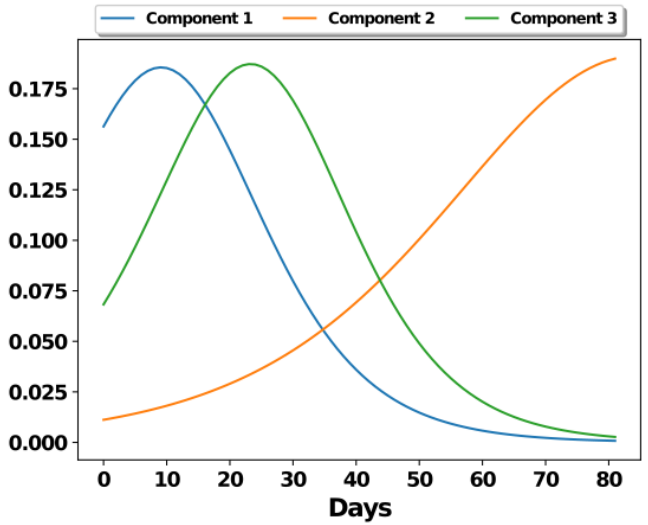
- Task
  - Case counts for different locations and signals over time.
  - Prediction of epidemic trends for all locations simultaneously.



Sneak Preview

- In this work we:
  - Propose STELAR, a data efficient tensor factorization method to predict the evolution of epidemic trends.
  - Perform experiments on real county- and state-level COVID-19 data.
  - Demonstrate superior prediction performance compared to baselines.
  - Identify interesting latent patterns of the epidemic.





Related Work

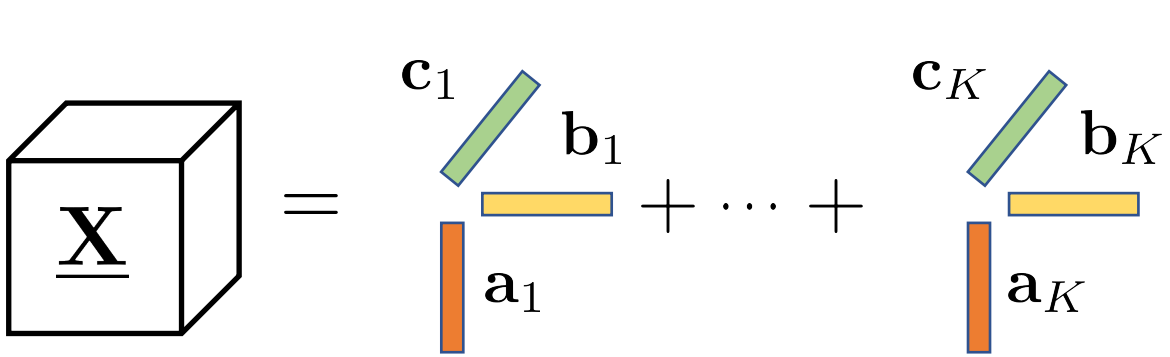
- Mechanistic Models
  - Susceptible-Infected-Recovered (SIR) model [Kermack and McKendrick, 1927].
  - Susceptible-Exposed-Infected-Recovered (SEIR) [Cooke and Van Den Driessche, 1996].
  - Rely on a system of differential equations.
  - Do not require much training data.
  - Restrictive, cannot leverage rich / “collaborative” information.
- Machine Learning Models
  - Time series prediction problem.
  - LSTM [Chimmula and Zhang 2020], [Yang et al. 2020].
  - GNN [Gao et al, 2020], [Kapoor et al. 2020].
  - Learn only from data.
  - Usually require large amount of training data.
- This work
  - Nonnegative tensor factorization (Canonical Polyadic Decomposition).
  - Latent epidemiological dynamics (SIR model) to capture common epidemic profile sub-types.
  - Learns from limited data and can extract interpretable latent components.

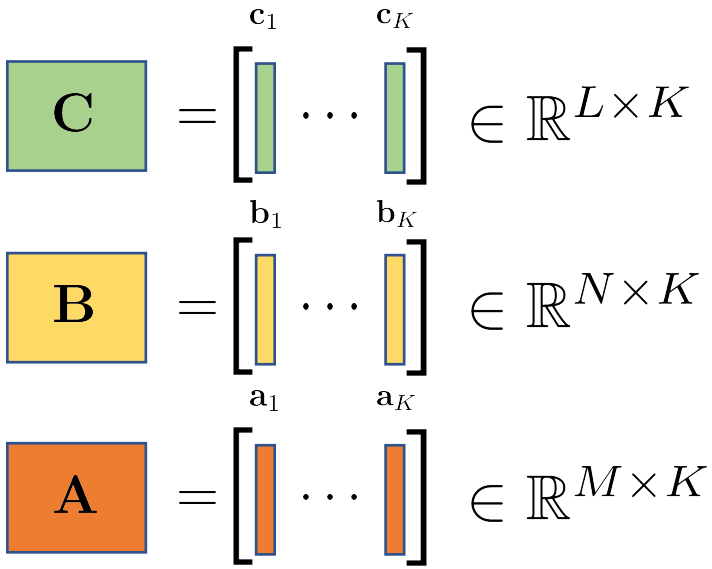
Canonical Polyadic Decomposition (CPD)

- A *3*-way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{M \times N \times L}$ admits a decomposition of rank *K* if it can be decomposed as a sum of *K* rank-1 tensors

$$\underline{\mathbf{X}} = \sum_{k=1}^{K} \mathbf{a}_k \circ \mathbf{b}_k \circ \mathbf{c}_k$$

$$\underline{\mathbf{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$





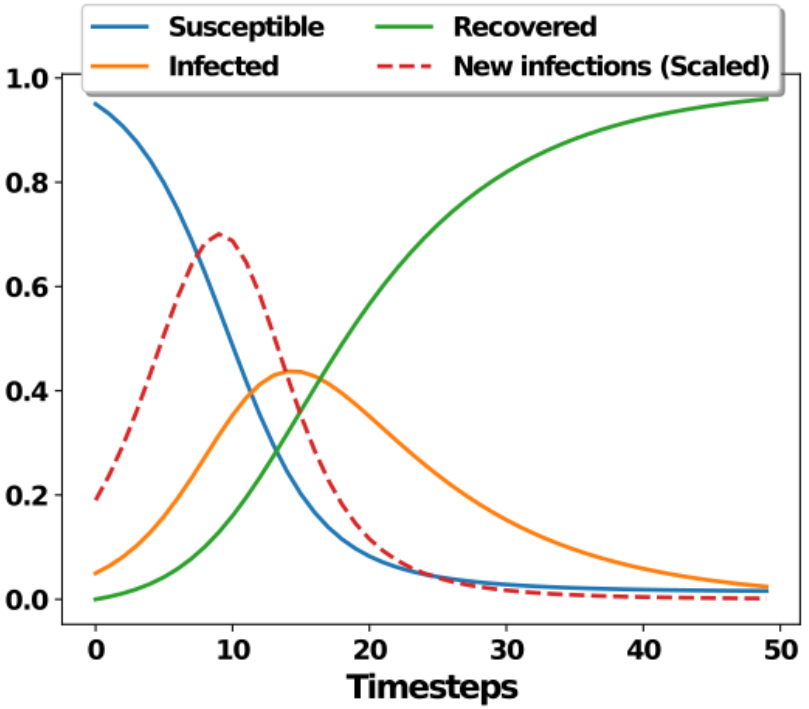
- Element-wise: $\underline{\mathbf{X}}(m,n,t) = \sum_{k=1}^{K} a_{m,k} b_{n,k} c_{t,k}.$
- Matrix unfolding: $\underline{\mathbf{X}}^{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T$
- Vector: $\mathrm{vec}(\underline{\mathbf{X}}) = (\mathbf{C} \odot \mathbf{B} \odot \mathbf{A})\mathbf{1}.$

The SIR Model

- SIR model [Kermack and McKendrick, 1927].
  - Susceptible S(t), infected I(t) and recovered R(t) subpopulations. N is the total population.

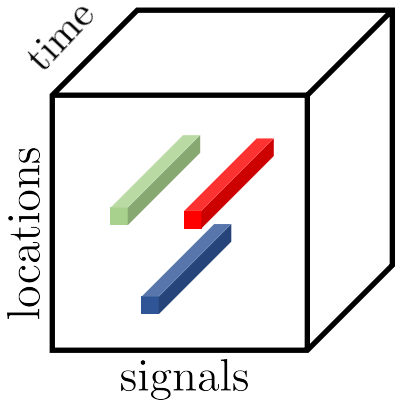
$$S(t) - S(t-1) = -\beta S(t-1)I(t-1)/N$$
$$I(t) - I(t-1) = \beta S(t-1)I(t-1)/N - \gamma I(t-1)$$
$$R(t) - R(t-1) = \gamma I(t-1)$$

  - β : rate of spread.
  - γ: recovery rate.



- New infections: $C(t) := \beta S(t-1)I(t-1)/N$
- In practice, we observe C(t) and want to find the model parameters.
- Example
  - $N = 1, S(0) = 0.95, I(0) = 0.05, R(0) = 0, \beta = 0.4, \gamma = 0.1.$

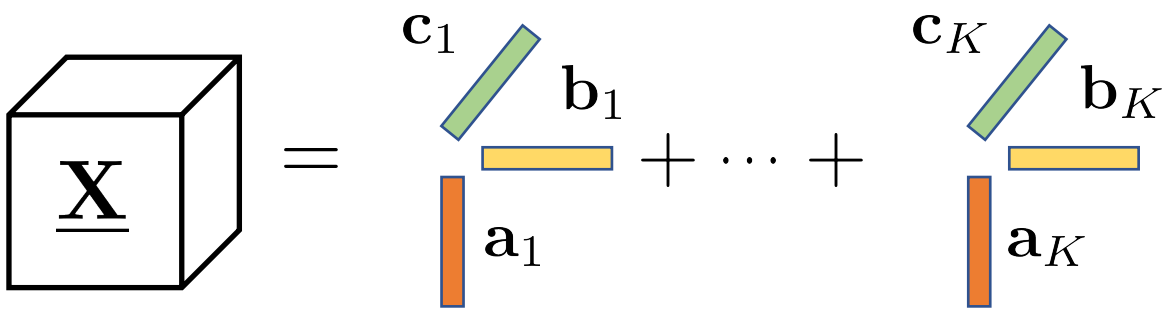
Key Idea



- M locations
  - Counties/states in the US.
- N signals
  - Daily new infections, ICU patients, hospitalized patients, etc.
- At time t, value of the nth signal at location m is denoted as $x_{m,n,t}$.
- The dataset can be naturally described by a 3-way spatio-temporal tensor $\underline{\mathbf{X}} \in \mathbb{R}^{M \times N \times L}$.

$$\underline{\mathbf{X}}(m, n, t) := x_{m,n,t}$$

- We would like to predict the frontal slabs $\underline{\mathbf{X}}(:, :, t)$ for L_o timesteps ahead.



Impose SIR constraints on the latent time factor!

$$c_{t,k} = \beta_k S_k(t-1) I_k(t-1)$$
$$S_k(t) = S_k(t-1) - \beta_k S_k(t-1) I_k(t-1)$$
$$I_k(t) = I_k(t-1) + \beta_k S_k(t-1) I_k(t-1) - \gamma_k I_k(t-1)$$
$$s_k = S_k(0), i_k = I_k(0)$$

Problem Formulation

$$\min_{\substack{\mathbf{A},\mathbf{B},\mathbf{C},\\ \boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{s},\mathbf{i}}} \|\underline{\mathbf{X}} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]\|_F^2 + \mu\left(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2\right) + \nu \sum_{k=1}^{K}\sum_{t=1}^{L}\left(c_{t,k} - \beta_k S_k(t-1)I_k(t-1)\right)^2$$

data fitting term — Frobenius norm regularization — SIR model regularization

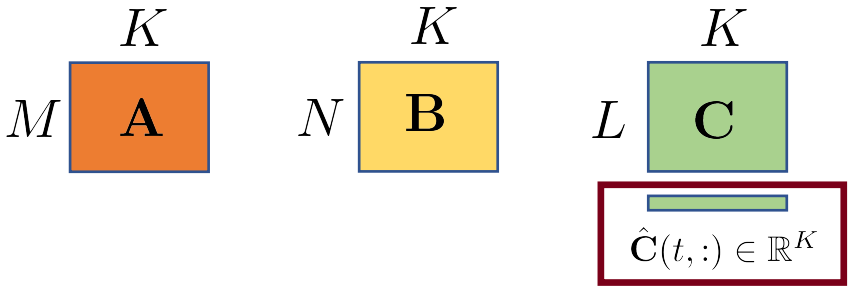
$$\text{s. t. } \boxed{\mathbf{A} \geq \mathbf{0}, \mathbf{B} \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0},}$$
$$\boldsymbol{\beta} \geq \mathbf{0}, \boldsymbol{\gamma}, \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}, \mathbf{i} \geq \mathbf{0}$$

$$\begin{aligned} &S_k(t) = S_k(t-1) - \beta_k S_k(t-1)I_k(t-1), \\ &I_k(t) = I_k(t-1) + \beta_k S_k(t-1)I_k(t-1) - \gamma_k I_k(t-1), \\ &s_k = S_k(0), i_k = I_k(0). \end{aligned}$$

Prediction

After the convergence of the optimization algorithm, we have estimates of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and parameters $\{\beta_1, \cdots, \beta_K\}$, $\{\gamma_1, \cdots, \gamma_K\}$, $\{s_1, \cdots, s_K\}$, $\{i_1, \cdots, i_K\}$.

$$\underline{\mathbf{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$



Let $\hat{\mathbf{C}}(t,:) \in \mathbb{R}^K$ be the prediction of the temporal information at a future time point t using estimates $\{\beta_1, \cdots, \beta_K\}$, $\{\gamma_1, \cdots, \gamma_K\}$, $\{s_1, \cdots, s_K\}$, $\{i_1, \cdots, i_K\}$.

We predict an entire "future" slab using $\widehat{\underline{\mathbf{X}}}(:,:,t) = \mathbf{A}\mathrm{diag}(\hat{\mathbf{C}}(\mathrm{t},:))\mathbf{B}^{\mathrm{T}}$.

Prediction

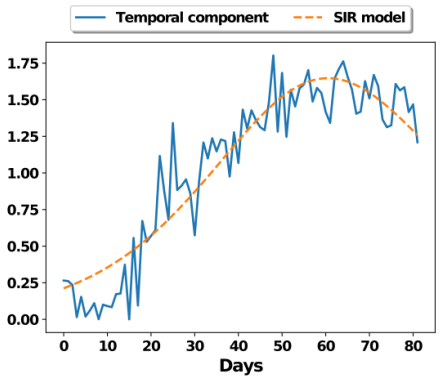
Element-wise: $\underline{\mathbf{X}}(m,n,t) = \sum_{k=1}^{K} a_{m,k} b_{n,k} c_{t,k} = \sum_{k=1}^{K} a_{m,k} b_{n,k} \beta_k S_k(t-1) I_k(t-1).$

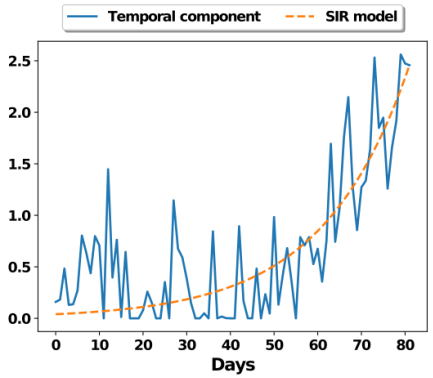
$\longrightarrow$ Our model expresses the evolution of a signal as **weighted sum of K separate SIR models**.

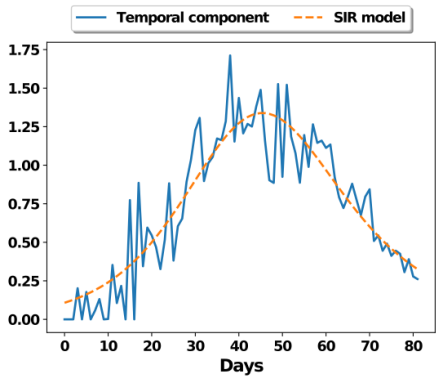
$\longrightarrow$ Captures **correlations** between different **locations and signals** through their latent representations.

$$L \overbrace{\underline{\mathbf{X}}^{(3)}}^{M \times N} = L \overbrace{\mathbf{C}}^{K} \times K \overbrace{(\mathbf{B} \odot \mathbf{A})^T}^{M \times N}$$

$[\underline{\mathbf{X}}(1,1,:), \underline{\mathbf{X}}(1,2,:), \ldots, \underline{\mathbf{X}}(M,N,:)]$







Optimization

$$\min_{\substack{\mathbf{A},\mathbf{B},\mathbf{C},\\ \boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{s},\mathbf{i}}} \|\underline{\mathbf{X}} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]\|_F^2 + \mu\left(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2\right) + \nu \sum_{k=1}^{K}\sum_{t=1}^{L}\left(c_{t,k} - \beta_k S_k(t-1)I_k(t-1)\right)^2$$

$$\begin{aligned}
\text{s. t. } \quad & \mathbf{A} \geq \mathbf{0}, \mathbf{B} \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0}, \\
& \boldsymbol{\beta} \geq \mathbf{0}, \boldsymbol{\gamma}, \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}, \mathbf{i} \geq \mathbf{0} \\
& S_k(t) = S_k(t-1) - \beta_k S_k(t-1) I_k(t-1), \\
& I_k(t) = I_k(t-1) + \beta_k S_k(t-1) I_k(t-1) - \gamma_k I_k(t-1), \\
& s_k = S_k(0), i_k = I_k(0).
\end{aligned}$$

To update factor matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and the SIR model parameters we rely on alternating optimization.

By fixing all variables except for $\mathbf{A}$ the resulting subproblem is a nonnegative least squares problem.

Similarly, for $\mathbf{B}$, $\mathbf{C}$.

For the SIR model parameters, we rely on a few projected gradient descent steps.

Experiments

- Dataset information
  - US county-level data from the Johns Hopkins University (JHU) [Dong, Du and Gardner, 2020]
  - Large patient claims dataset from IQVIA.

  - Daily counts of 12 International Classification of Diseases ICD-10 codes observed in each county.
  - Current Procedural Terminology (CPT) codes related to hospitalization and utilization of intensive care unit (ICU).

  - The total number of counties was 133.
  - The total number of signals was 15.
  - The time window was from 03-24-2020 to 06-26-2020 (95 days).

- First experiment : 85 days used for training, 10 days for test.
- Second experiment : 80 days used for training, 15 days for test.

- Metrics: RMSE, MAE.

Results – County-level Prediction

New infections

Model	$L_o = 10$ RMSE	$L_o = 10$ MAE	$L_o = 15$ RMSE	$L_o = 15$ MAE
Mean	304.1	122.0	269.5	108.5
SIR	156.2	62.2	159.1	63.6
SEIR	177.1	72.9	163.2	69.7
LSTM (w/o feat.)	203.6	77.1	191.0	81.7
LSTM (w/ feat.)	162.3	68.2	187.6	78.3
STAN	164.2	61.1	152.6	61.8
STELAR ($\nu = 0$)	149.2	61.5	152.8	66.9
STELAR	**127.5**	**55.6**	**136.1**	**61.7**

Hospitalized patients

Model	$L_o = 10$ RMSE	$L_o = 10$ MAE	$L_o = 15$ RMSE	$L_o = 15$ MAE
Mean	125.0	77.0	123.3	77.1
SIR	46.5	27.2	48.7	27.7
SEIR	39.1	23.9	41.1	25.7
LSTM (w/o feat.)	45.6	23.6	54.8	31.2
LSTM (w/ feat.)	42.5	23.3	47.5	26.8
STAN	30.6	17.3	42.8	24.2
STELAR ($\nu = 0$)	28.6	16.6	46.8	21.1
STELAR	**24.0**	**15.1**	**36.0**	**18.0**

- New infections
  - $L_o = 10$: 18% lower RMSE and 9% lower MAE compared to the best performing baselines.
  - $L_o = 15$: 10% lower RMSE and the same MAE compared to the STAN model.
- Hospitalized patients
  - $L_o = 10$: 21% lower RMSE and 12% lower MAE compared to STAN.
  - $L_o = 15$: 12% lower RMSE and 25% lower MAE compared to the best baselines.

Results – State-level Prediction

New infections

Model	$L_o = 10$ RMSE	$L_o = 10$ MAE	$L_o = 15$ RMSE	$L_o = 15$ MAE
Mean	309.0	258.7	325.8	273.1
SIR	186.1	133.8	186.9	134.5
SEIR	162.4	127.0	162.6	130.2
LSTM (w/o feat.)	187.5	138.1	419.7	356.0
LSTM (w/ feat.)	197.9	151.6	359.2	286.5
STAN	**74.1**	**60.1**	**100.5**	79.6
STELAR ($\nu = 0$)	140.8	104.0	127.8	95.0
STELAR	117.8	89.8	107.3	**79.4**

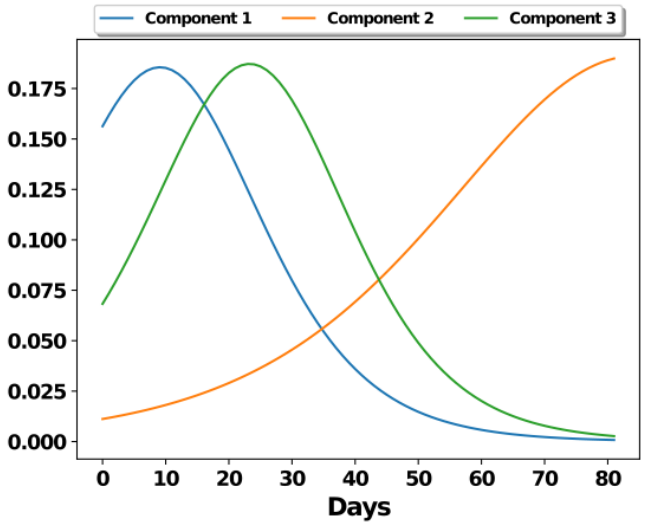
Hospitalized patients

Model	$L_o = 10$ RMSE	$L_o = 10$ MAE	$L_o = 15$ RMSE	$L_o = 15$ MAE
Mean	685.5	553.9	729.2	586.0
SIR	343.4	252.4	367.8	266.0
SEIR	109.0	97.6	192.5	154.3
LSTM (w/o feat.)	280.9	187.5	416.1	308.7
LSTM (w/ feat.)	295.3	182.3	276.0	208.5
STAN	100.3	73.6	177.7	144.7
STELAR ($\nu = 0$)	118.3	84.4	126.6	**75.2**
STELAR	**56.8**	**43.9**	**113.6**	83.8

In all cases except one, joint optimization and SIR model fitting improves the performance.

Results – Interpretability

We trained our model using K= 30.



The strongest 3 temporal components of our model.

Component 1	Component 2	Component 3
New York (NY)	L.A (CA)	Nassau (NY)
Westchester (NY)	Cook (IL)	L.A (CA)
Nassau (NY)	Milwaukee (WI)	Essex (NJ)
Bergen (NJ)	Fairfax (VA)	Wayne (MI)
Miami-Dade (FL)	Hennepin (MN)	Oakland (MI)
Hudson (NJ)	Montg. (MD)	Middlesex (NJ)
Union (NJ)	P. George's (MD)	New York (NY)
Phila. (PA)	Dallas (TX)	Phila. (PA)
Passaic (NJ)	Orange (CA)	Cook (IL)
Essex (NJ)	Harris (TX)	Bergen (NJ)

Counties that contribute more to each of the strongest 3 rank-1 components.

Component 1	Component 2	Component 3
New infections	New infections	Hosp. patients
Hosp. patients	Hosp. patients	ICU patients
ICU patients	ICU patients	J96
J96	J96	N17
R09	R05	R06

Signals that contribute more to each of the strongest 3 rank-1 components.

J96--Respiratory failure, N17--Acute kidney failure, R05--Cough, R06--Abnormalities of breathing, R09--Other symptoms and signs involving the circulatory and respiratory system.

Concluding Remarks

- We proposed STELAR – a data efficient and interpretable method based on constrained nonnegative tensor factorization.

- Our method enables long-term prediction of future slabs by incorporating latent epidemiological regularization.

- We demonstrated the ability of our method to make accurate predictions on real COVID-19 data.

THANK YOU!

Questions?