Spatio-Temporal PRRS Epidemic Forecasting via Factorized Deep Generative Modeling

Mohammadsadegh Shamsabardeh, Bahar Azari, Beatriz Martínez-López
University of California Davis

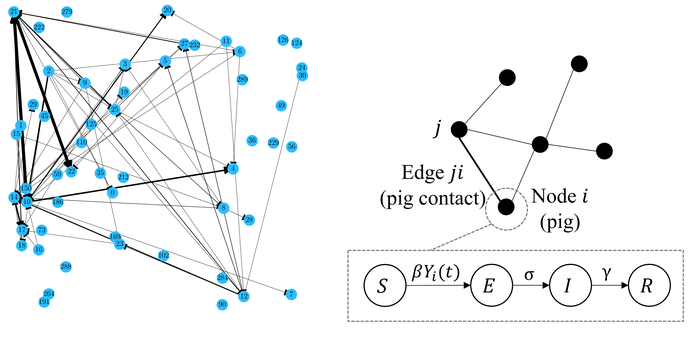
icassp 2022 Singapore

1. Motivation

- The Porcine Reproductive and Respiratory Syndrome (PRRS) is arguably the most challenging and costly viral infectious disease in the US swine industry [1].
- Usually, these data do not satisfy the granularity required for learning an advanced predictive model (e.g. diagnostic samples are only taken once or twice per month per farm).
- However, using the real-world data, we can simulate epidemics to produce fine-grained time series data to analyze it further with an advanced novel prediction method based on a generative and variational inference model.

2. Time Series Data Simulation

- Based on the rich database of an extensive anonymous swine production system located in the Midwest of the United States, we have access to farm-level pig shipment data, and PRRSV testing results [2].
- From 2006 to 2021, there have been over 260,000 movement records to or from farms within this production system.
- For each movement entry, the data includes the source and destination information, the number of transported pigs, and the date of the movement.
- Based on the farm-level shipment data we generate a *farm-level movement network* for the entire production system. Furthermore, the frequent PRRSV testing in each farm gives insight into how the virus is transmitted, e.g., what is the virus's transmission rate, incubation time, etc. Using the SEIR model we can produce a *pig-level contact network*, [3].



(a) farm-level (b) pig-level

Contact Network. (a) The swine shipment network (directed graph). The premises are displayed by a number-labeled node and edge weights corresponds to the shipment rate. The between-premises shipment rate network is showcased for 10% of nodes randomly selected among over 300 existing nodes. (b) **Top:** Pig level network graph. **Bottom:** State-transition diagram for a single node.

3. Farm Disease Propagation Prediction

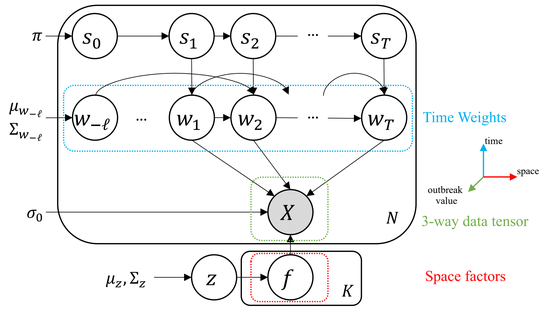
- Our spatio-temporal data indicates the number of pigs categorized within a particular stage, e.g., infected, recovered, etc., in every time instance in each farm. We denote this data as the matrix $X \in \mathbb{R}^{T \times D}$, where T is the number of time points and D the number of spatial locations, e.g., the number of farms.
- Building on previous work by [4], our assumption is that X can be decomposed into a weighted summation of $K \ll D$ factors over time as:

$$X \approx [w_1, \cdots, w_T]^\top [f_1; \cdots; f_K] = W^\top F, \tag{1}$$

  where $f_k \in \mathbb{R}^D$ is the k^{th} spatial factor and $w_t \in \mathbb{R}^K$ is the weight vector at time t.
- Our intuition for adopting this model for some pig-specific collected measurements in D farm over T time points is that there are $K \ll D$ underlying factors using which we can approximate the overall dynamics of the disease propagation in the data.
- We assume that the weights, $W = \{w_t\}_{t=1}^T$, are generated according to a set of temporal lags, ℓ, through a deep probabilistic switching auto-regressive model. These weights are furthermore governed by a Markovian chain of discrete latent states, $\mathcal{S} = \{s_t\}_{t=1}^T$ as follows: $w_t \sim p(w_t | w_{t-\ell}, s_t)$, $s_t \sim p(s_t | s_{t-1})$. In addition, we assume that spatial factors, $F = \{f_k\}_{k=1}^K$, are controlled by a shared low dimensional latent variable, z, as follows: $f_{1:K} \sim p(F|z)$, $z \sim p(z)$.
- We train the model using stochastic variational methods by approximating the posterior $p_\theta(\mathcal{S}, W, z, F | X)$ using a variational distribution $q_\phi(\mathcal{S}, W, z, F)$, and by maximizing a lower bound (known as ELBO) $\mathcal{L}(\theta, \phi) \leq \log p_\theta(X)$:

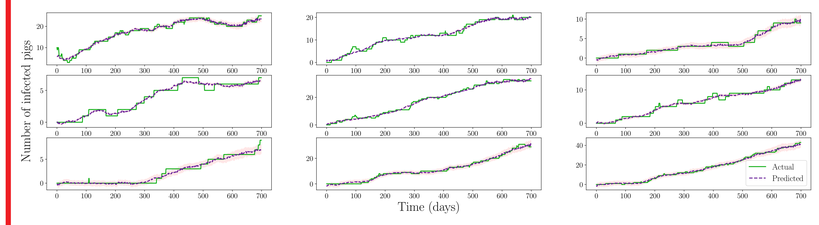
$$\begin{aligned} \mathcal{L}(\theta,\phi) &= \mathbb{E}_{q_\phi(\mathcal{S},W,z,F)} \left[\log \frac{p_\theta(X,\mathcal{S},W,z,F)}{q_\phi(\mathcal{S},W,z,F)} \right] \\ &= \log p_\theta(X) - \mathrm{KL}(q_\phi(\mathcal{S},W,z,F) \,||\, p_\theta(\mathcal{S},W,z,F|X)). \end{aligned} \tag{2}$$

By maximizing the bound with respect to the parameters θ, we learn the generative distribution over datasets $p_\theta(X)$, and by maximizing the bound over the parameters ϕ, we do Bayesian inference by approximating the distribution $q_\phi(\mathcal{S}, W, z, F) \simeq p_\theta(\mathcal{S}, W, z, F | X)$ over latent variables for each data point.



4. Experimental Results

- We used time series of epidemic progression from over 300 farms simulated for 700 time points.
- We kept last 20% of the time series as the test set.
- We then performed a short-term prediction tasks by adopting a rolling prediction scheme reported in [5].
- For short-term prediction, the next time point is predicted using the generative model and spatial factors learned on the train set.
- We reported the test set normalized root-mean-square error (NRMSE%), which is related to the expected negative test-set log-likelihood for the case of Gaussian distributions, and it is used for evaluating the predictive generative models.
- We obtained NRMSE of 2.5% averaged over all the farms.
- The figure below shows the number of infected pigs over time for nine selected farms.



Short-term (one-day) prediction. Each plot demonstrates the actual number of infected pigs in the simulated data (solid green), the mean estimate of the predictive model (dashed purple), and the standard deviation of the prediction estimate (shaded red error bar). Each row represents neighbouring farms in movement network.

Acknowledgement

This project was partially funded by the NSF BIGDATA:IA Award #1838207 and NSF Track-D award #2134901. Authors would like to acknowledge swine industry collaborators for the provision of data.

References

[1] Holtkamp et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on united states pork producers. *Journal of Swine Health and Production*, 21(2):72–84, 2013.

[2] M Shamsabardeh, Shabaz Rezaei, Jose Pablo Gomez, Beatriz Martínez-López, and Xin Liu. A novel way to predict prrs outbreaks in the swine industry using multiple spatio-temporal features and machine learning approaches. *Frontiers in Veterinary Science*, 6, 2019.

[3] Ferdousi et al. Generation of swine movement network and analysis of efficient mitigation strategies for african swine fever virus. *PloS one*, 14(12), 2019.

[4] Farnoosh et al. Deep switching auto-regressive factorization: Application to time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7394–7403, 2021.

[5] Chen et al. Missing traffic data imputation and pattern discovery with a bayesian augmented tensor factorization model. *Transportation Research Part C: Emerging Technologies*, 104:66–77, 2019.