CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting (Supplementary Material)

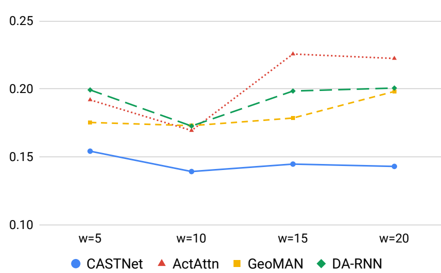
Ali Mert Ertugrul[1,2], Yu-Ru Lin[1], and Tugba Taskaya-Temizel[2]

[1] University of Pittsburgh, Pittsburgh PA, USA {ertugrul,yurulin}@pitt.edu
[2] Middle East Technical University, Ankara, Turkey ttemizel@metu.edu.tr

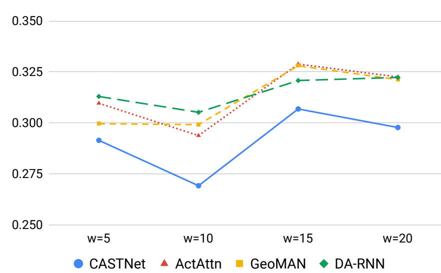
A Appendix

A.1 Window Size Analysis

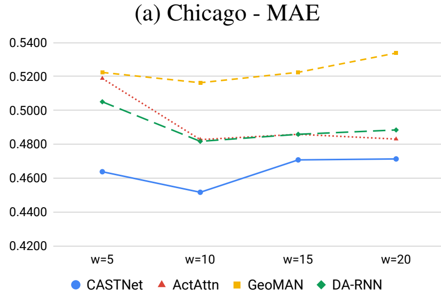
For RNN-based methods, we conducted experiments with different length of window size $w \in \{5, 10, 15, 20\}$, and shared the results for $w = 10$ in the paper, which is the best setting for all models. We also provide the performance results of those methods (ActAttn [1], GeoMAN [2] and DA-RNN [3]) as well as our method CASTNet, for different window size w settings in Fig. 1. Based on the results, while our proposed method CASTNet achieves superior performance in terms of MAE for all window size settings on both datasets, it significantly performs better than the baseline methods in terms of RMSE at almost all cases for both datasets.



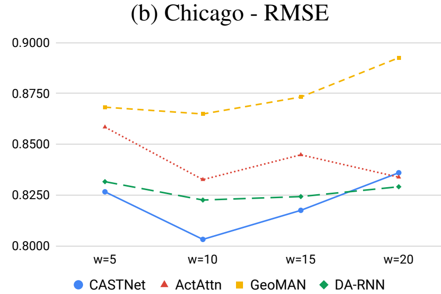
(a) Chicago - MAE



(b) Chicago - RMSE

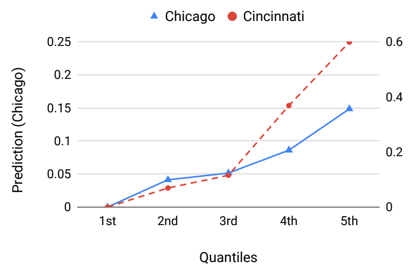


(c) Cincinnati - MAE

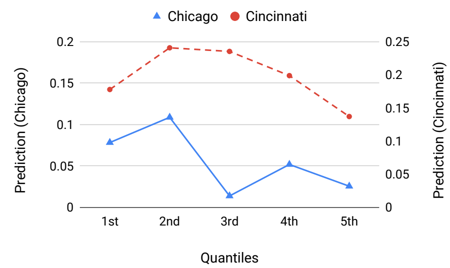


(d) Cincinnati - RMSE

Fig. 1: **MAE and RMSE results w.r.t different window size w.**



(a) Population



(b) Gender Diversity

Fig. 2: **Median prediction results for each quantile.**

A.2 Static Feature Analysis

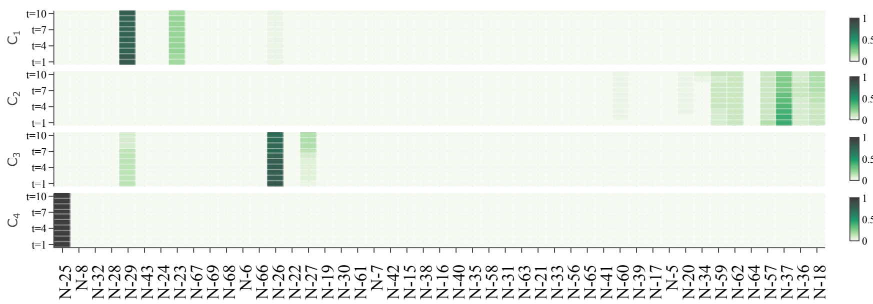
In order to analyze the effect of static features (one-by-one) on the prediction, we performed additional experiments. We divided the neighborhoods into 5-quantiles based on values of corresponding static feature. In Fig. 2, we show the median forecasting results on the test set for each of these 5 groups with respect to the features *population* and *gender diversity*. We observe that, our model predicts larger opioid overdoses for neighborhoods having larger population (Fig. 2a) for both cases. Also, it predicts larger opioid overdose for the neighborhoods with lower and moderate gender diversity in Cincinnati and neighborhoods with lower gender diversity in Chicago (Fig. 2b). However, these findings require further investigation for validation with statistical tests.

A.3 Community Memberships

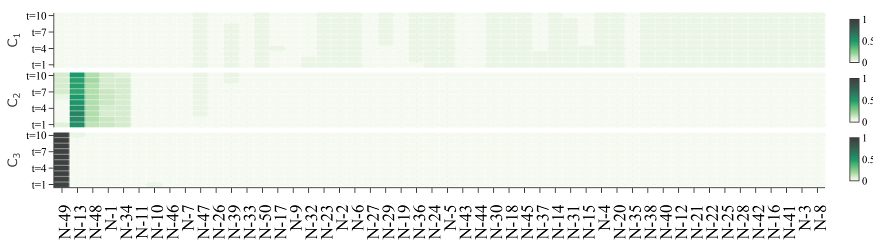
CASTNet learns different representation subspaces (communities) of global dynamics unlike the previous work [2, 1], and each community consists of a group of different members due to orthogonality penalty. To analyze these communities and to see how these communities differ from each other, we visualize the average spatial attention weights over time on test set, which represents the community memberships of the neighborhoods as shown in Fig. 3. Note that the neighborhoods are ordered by the number of crime incidents. We observe that most of the locations have dedicated to one community for both Chicago and Cincinnati (Fig. 3a and Fig. 3b). This indicates that each community learned a different representation subspace of the global dynamic features and there is no redundant community.

References

1. Ertugrul, A.M., Lin, Y.R., Chung, W.T., Yan, M., Li, A.: Activism via attention: interpretable spatiotemporal learning to forecast protest activities. EPJ Data Science **8**(1), 5 (2019)
2. Liang, Y., Ke, S., Zhang, J., Yi, X., Zheng, Y.: Geoman: Multi-level attention networks for geo-sensory time series prediction. In: IJCAI. pp. 3428–3434 (2018)
3. Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., Cottrell, G.: A dual-stage attention-based recurrent neural network for time series prediction. arXiv preprint arXiv:1704.02971 (2017)



(a) Chicago



(b) Cincinnati

Fig. 3: **Spatial attention weights in each community.** It represents the average spatial attention weights over time for each location in the corresponding community, which is called community membership. Where x-axis represents the formal location identifiers (community areas for Chicago and SNAs for Cincinnati), y-axis indicates the time steps.