HintNet: Hierarchical Knowledge Transfer Networks for Traffic Accident Forecasting on Heterogeneous Spatio-Temporal Data
Supplementary Document

Bang An* Amin Vahedian† Xun Zhou* W. Nick Street* Yanhua Li‡

1 Detailed Feature List

Temporal Feature F_T: the day of the week, day of the year for time, whether this is a holiday, whether this is weekend, month of the year.

Spatial Features F_S: (1) 13 counts of different POI categories: eat-drink, going-out, sights-museums, transport, accommodation, shopping, leisure-outdoor, administrative-areas-buildings, natural-geographical ,petrol-station, atm-bank-exchange, toilet-rest-area, and hospital-health-care-facility. (2) 6 features indicating basic road conditions: Annual Average Daily Traffic(AADT), average traffic speed, average mileage for each road, number of intersections, the total mileage of road system, and total annually traffic volume. (3) 10 spectral features as proposed by Yuan et al. [1].

Definition 1.1. Spatial-Temporal Feature is defined as F_{ST} (1) 9 counts of different weather features: consist of, average air temperature, highest temperature, lowest temperature, wind speed, precipitation, snowfall, snow depth, dew point temperature, and MERRA[1]. (2) 4 counts of the traffic condition feature includes average traffic speed, normal vehicle traffic volume, truck traffic volume, and Occupancy. The detailed features are listed in Table 2 and Table 3.

Visualization: Figure 2 (a) illustrates the traffic accident distribution from 2016 to 2018. Figure 2 (b) shows the road network in the state of Iowa. Figure 2 (c) demonstrates the locations of observation stations.

2 Solution

A simple example of M-RSP: Figure 1 is a simple example of M-RSP on a 5×5 grid with min_neighbors $\beta = 3$, 7, and 9 respectively and $\epsilon = 1$. The numbers in each cell represent the accident counts, and grey cells all have 0. In each figure, yellow, orange, red grid cells represent the partitioned region on each level. When $\beta = 3$, RSP partitioned the whole grid into grey and

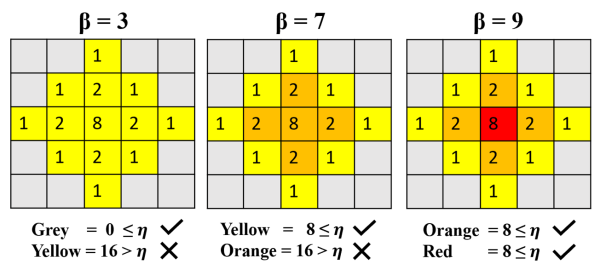


Figure 1: **Illustration of Multi-level Partitioning** ($\eta = 8$). Each color represents a partitioned group. The numbers represent accident counts in each cell. Last image

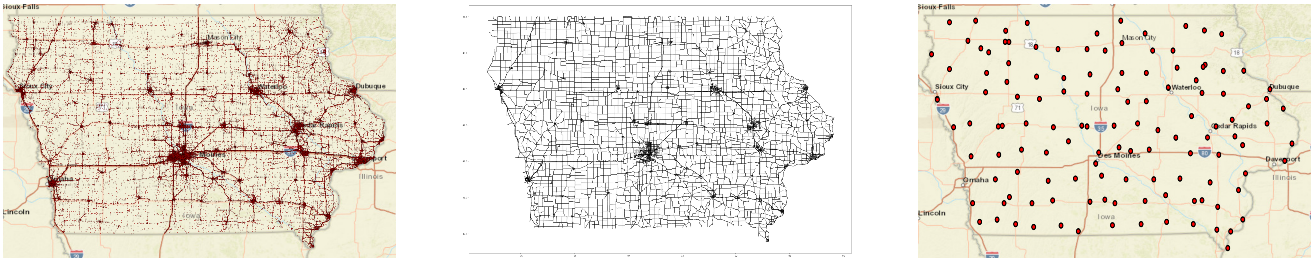
yellow. For both of them, M-RSP compares the number of accidents in each region with η. In this case, the yellow region is risky and will be further partitioned into two levels. The grey region is less-risky, so grey cells are assigned with a level label that equals β. In the next level, with higher β, RSP splits the previous yellow region into an orange and a smaller yellow region. The accident counts in each region are checked with hη again. Orange cells are risky, and yellow region is less-risky. The same process repeats until reaching maximum β. In the last level, the remaining cells are all assigned with maximum β, because they represent the highest-risk regions. The threshold η is a tunable hyper-parameter.

3 Evaluation

Implementation Details: The proposed model is trained by minimizing the mean square error loss using Adam optimizer with setting $\alpha = 0.0001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We design an early stop mechanism that model terminates training when the best validating loss stops decreasing for 5 epochs. For the parameter setting of Multi-level partitioning, we set querying range $\epsilon = 1$, $\theta = 25$, $\eta = 500$, and $\lambda = 10$. The sub-region size w is set as 5×5 or 7×7 depends on the size of the grid. We test setting granularity K equals 1, 2, and 3. The experiment of choosing appropriate granularity is presented in the ablation study part. To

*University of Iowa, {bangan, xunzhou, nickstreet}@uiowa.edu
†Northern Illinois University, avahediankhezerlou@niu.edu
‡Worcester Polytechnic Institute, yli15@wpi.edu

[1] https://mesonet.agron.iastate.edu/request/coop/obs-fe.phtml



(a) traffic accidents in Iowa (b) Iowa Road Network COOP Observation Stations

Figure 2: Visualization of traffic accidents, road network, and COOP

Table 1: Number of parameters comparison (in thousands)

	ConvLSTM	**GSNet**	**Hetero-ConvLSTM**	**HintNet**
16×16	$3,088$	$1,106$	$1,112$	$1,182$
32×32	$3,088$	$16,835$	$3,088$	$1,182$
64×64	$6,052$	$268,497$	$27,792$	$2,277$
128×64	$6,052$	$536,994$	$64,849$	$2,277$

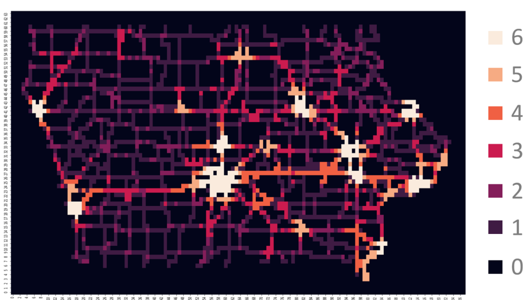


Figure 3: Multi-level partitioned result in the area of Iowa with size 128×64. Higher cluster number represents higher accident risk. Level 0 is noise.

achieve the best performance and reasonable training efficiency, we set k as 2. In this case, the original partitioned levels are merged two levels by two levels. Level 0 is the noise level filtered from the mask map. Cells at this level are not considered in the training phase and prediction phase. We have 6 partitioned levels eventually. Figure 3 shows the final outcome from the M-RSP algorithm represents the accident risk level in the study area. The brighter cells with dense urban road systems indicate higher accident risk. On the opposite, the darker cells with lower accident risk are more likely to be in rural areas.

GSNet The samples weights $\lambda_i \in I$ of loss function are set to 0.05, 0.2, 0.25, 0.5 when the corresponding number of accidents within the range (0), (1,2), (3,4), and (greater than 4), respectively.

Model Complexity: Table 1 shows the total number of parameters trained in the three major deep learning methods. With larger grids, the number of parameters in Hetero-ConvLSTM and GSNet surges dramatically due to the ensemble structure and graph convolution respectively. On the opposite, HintNet remains with similar complexity, because the graph convolution layers in HintNet are applied only on subregions. Therefore, HintNet can demonstrate a stable performance on areas with different size.

References

[1] Yuan et al. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *ACM SIGKDD*, pages 984–992, 2018.

Table 2: Feature Table

Feature Group	**Feature List**
Temporal Features F_T^i	5 calendar features
Spatial Features F_S^i	13 POI features, 6 basic road condition features, 10 SpatialGraph features
Spatio-Temporal Features F_{ST}^i	9 weather features and 4 real-time traffic condition features

Table 3: Feature Table

Feature Group	**Feature List**
Temporal Features F_T^i	**5 calendar features:** day of the week, day of the year, month of the year, whether this is a holiday, whether this is weekend
Spatial Features F_S^i	**13 POI features:** eat-drink, going-out, sights-museums, transport, accommodation, shopping, leisure-outdoor, administrative-areas-buildings, natural-geographica **6 basic road condition features:** Annual Average Daily Traffic(AADT), average traffic speed, average mileage for each road, number of intersections, the total mileage of road system, and total annually traffic volume **10 SpatialGraph features**
Spatio-Temporal Features F_{ST}^i	**9 weather features:** average air temperature, highest temperature, lowest temperature, wind speed, precipitation, snowfall, snow depth, dew point temperature, and MERRA **4 real-time traffic condition features:** average traffic speed, normal vehicle traffic volume, truck traffic volume, and Occupancy