Travel Speed Forecast using Continuous Conditional Random Fields

Nemanja Djuric, Vladan Radosavljevic, Vladimir Coric, Slobodan Vucetic

Department of Computer and Information Sciences, Temple University

Abstract

CCRF is a flexible, probabilistic framework that can seamlessly incorporate multiple traffic predictors and can exploit spatial and temporal correlations inherently present in traffic data. In addition to improving the prediction accuracy, the probabilistic approach also provides information about prediction uncertainty. Moreover, information about how important particular predictor and spatial-temporal correlations are can be easily extracted from the model. CCRF are also fault tolerant and can provide predictions even when some of the observations are missing. We applied several CCRF models on the problem of travel speed prediction in a range between 10 and 60 minutes ahead, and evaluated them on loop detector data from a 5.71-mile section of the I-35W highway in Minneapolis, MN. When compared to the linear regression models, the Mean Absolute Error was reduced by around 4%.

Continuous CRF model

CCRF is a log-linear model, and the conditional probability of output **y** (travel speed) given existing traffic state **x** is:

$$P(\mathbf{y}\,|\,\mathbf{x}) = \frac{1}{Z(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta})}\exp\Big(\sum_{i=1}^{N} A(\boldsymbol{\alpha}, y_i, \mathbf{x}) + \sum_{i\sim j} I(\boldsymbol{\beta}, y_i, y_j, \mathbf{x})\Big)$$

where the *Association* and *Interaction* potentials are defined as:

$$A(\boldsymbol{\alpha}, y_i, \mathbf{x}) = -\sum_{m=1}^{M} \delta_{mi}(\mathbf{x})\alpha_m (y_i - \theta_{mi}(\mathbf{x}))^2$$

$$I(\boldsymbol{\beta}, y_i, y_j) = -\sum_{k=1}^{K} w_{ij}^{k}\beta_k (y_i - y_j)^2$$

and Z(**x**, **α**, **β**) is a normalization function, and can be difficult to calculate:

$$Z(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = \int_{\mathbf{y}} \exp\Big(\sum_{i=1}^{N} A(\boldsymbol{\alpha}, y_i, \mathbf{x}) + \sum_{i\sim j} I(\boldsymbol{\beta}, y_i, y_j, \mathbf{x})\Big)d\mathbf{y}$$

Since potentials are linear combinations of quadratic functions of outputs **y**, the distribution corresponds to multivariate Gaussian distribution:

$$P(\mathbf{y}\,|\,\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\,|\boldsymbol{\Sigma}(\mathbf{x})|^{1/2}}\exp\Big(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x}))^T\,\boldsymbol{\Sigma}(\mathbf{x})^{-1}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x}))\Big)$$

Training phase: The task is to obtain the weights **α** and **β**. Training is performed by maximizing the log-likelihood using simple gradient ascent technique.

$$L(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{t=1}^{T} \log P(\mathbf{y}_t\,|\,\mathbf{x}_t)$$

$$(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\max}(L(\boldsymbol{\alpha},\boldsymbol{\beta})).$$

Prediction phase: Once we obtained weights and given the observations, the predicted speeds and uncertainties are just the parameters of the multivariate Gaussian distribution:

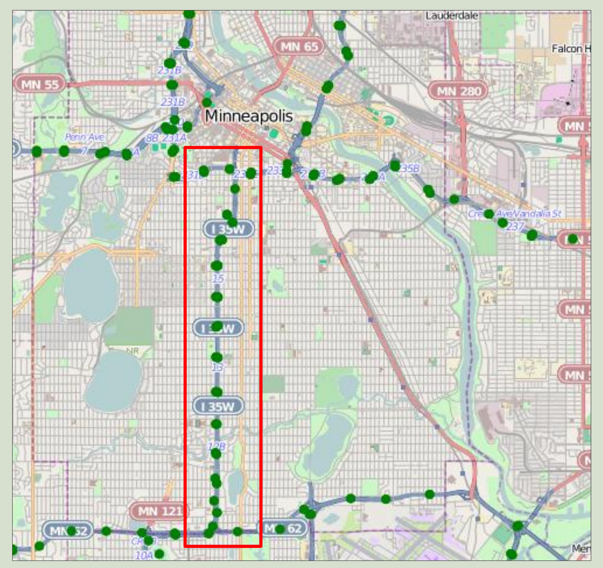
$$\hat{\mathbf{y}} = \boldsymbol{\mu} = \boldsymbol{\Sigma}\mathbf{b} \qquad \mathrm{var}(\mathbf{y}\,|\,\mathbf{x}) = diag(\boldsymbol{\Sigma})$$

where: $b_i = 2\sum_{m=1}^{M}\alpha_{mi}\delta_{mi}(\mathbf{x})\theta_{mi}(\mathbf{x})$, $\boldsymbol{\Sigma}^{-1} = 2(\mathbf{Q}_1 + \mathbf{Q}_2)$

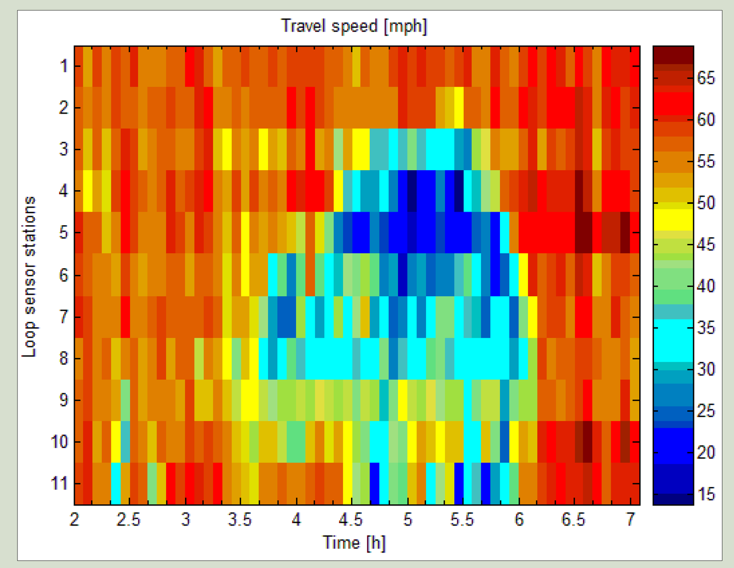
$$\mathbf{Q}_{1ij} = \begin{cases}\sum_{m=1}^{M}\alpha_{mi}\delta_{mi}(\mathbf{x}), i = j \\ 0, \text{otherwise}\end{cases} \qquad \mathbf{Q}_{2ij} = \begin{cases}\sum_{j=1}^{N}\sum_{k=1}^{K}\beta_k w_{ij}^{k}, i = j \\ -\sum_{k=1}^{K}\beta_k w_{ij}^{k}, i \neq j\end{cases}$$

Travel speed forecasting – Overview

We are predicting travel speeds up to 1h in future in 10-minute increments on 11 consecutive single-loop sensors of I-35W in Minneapolis, MN. The predictions were made for the time period from April 1st to July 1st, 2003, in daily interval from 14h to 19h.



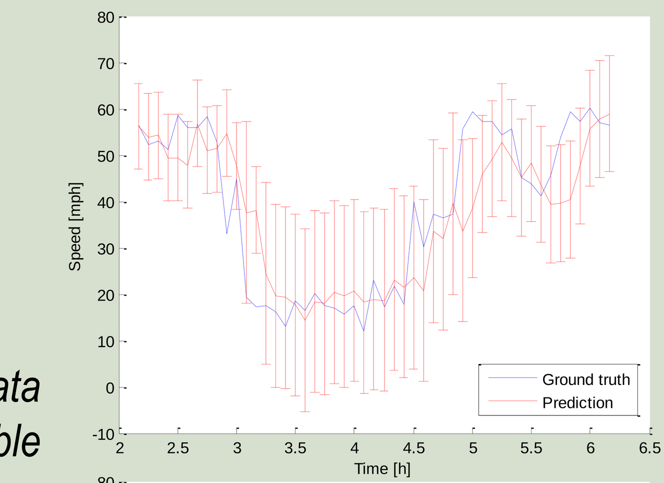
Considered road section (source: *OpenStreetMap*)



Typical day with a sizeable rush-hour period

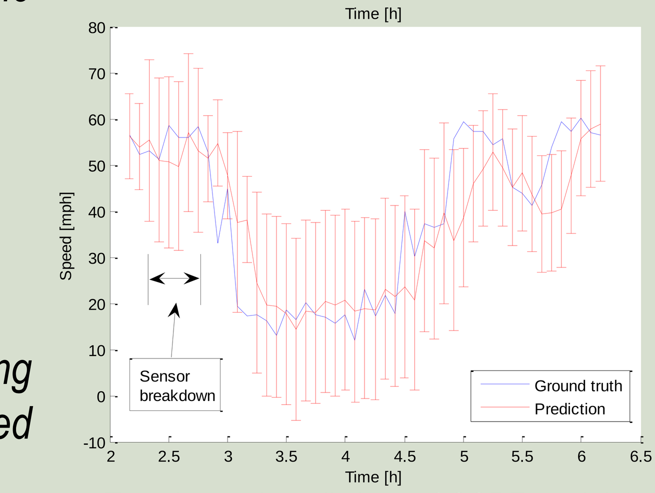
The model is very robust and can output predictions even when some data are missing.

Complete data available



However, this results in increased prediction uncertainty:

Some data missing or corrupted



Travel speed predictions of CCRF model 4

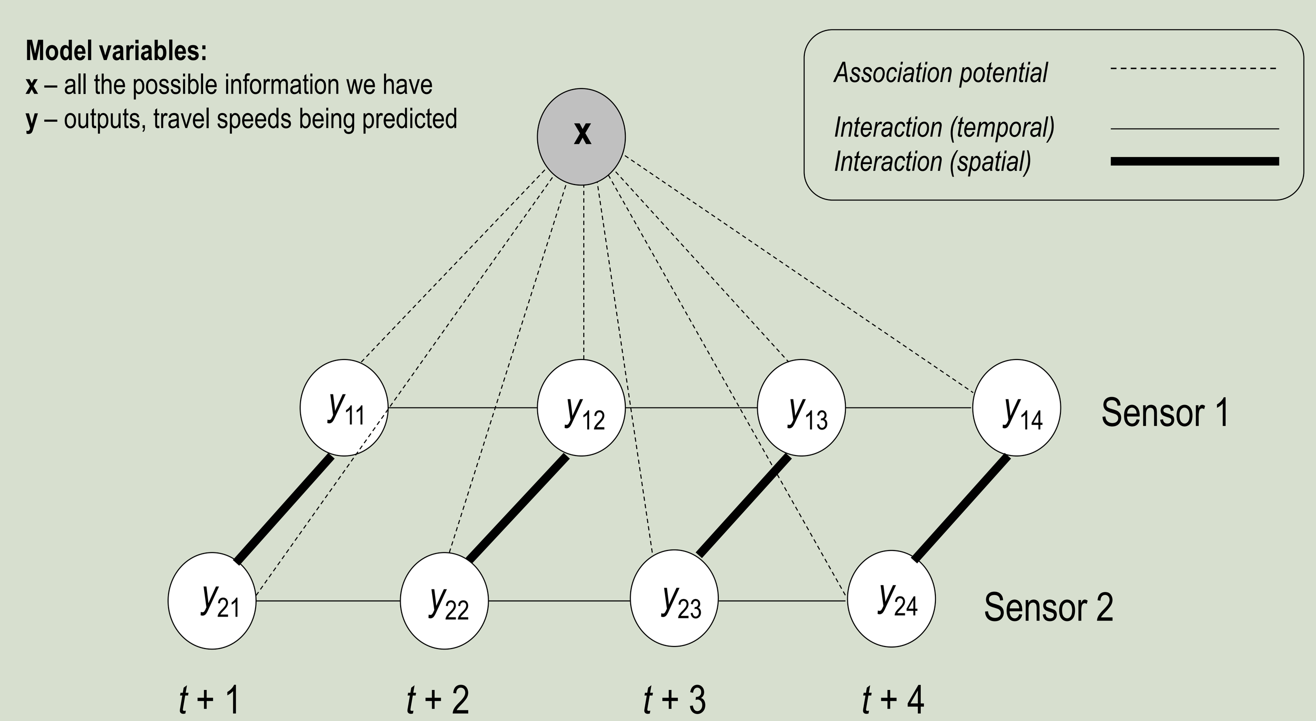
Continuous CRF – Travel speed predictions model

The first step of design is defining ASSOCIATION POTENTIAL:

- Encompasses all available predictors (Random Walk, Historical Mean, Neural Networks, …)
- Each predictor is assigned a weight (**α**)
- Weights tell us how much to believe to predictor, and are learned during training
- The higher the weight, more influence the predictor has on the final prediction

The second step of design is defining INTERACTION POTENTIAL:

- Includes all defined correlations
- Each correlation (neighborhood definition) assigned a weight (**β**)
- Weights tell us how strong are the defined correlations in the actual data
- The higher is the weight, the more influence do neighbors exert on each other



CCRF is easily expanded by adding new predictors, new neighborhood relations, and including different regimes (free-flow/congested). We included following simple predictors:

- Random Walk – predicts that current speed will not change
- Historical median – predicts that the speed will be a historical median for the section in question)
- Upstream state – predicts that the speed will equal the speed at the neighboring upstream section
- Downstream state – predicts that the speed will equal the speed at the neighboring downstream section

We tested several increasingly complex CCRF models:

- **Model 1**: 2 simple predictors, no interactions
- **Model 2**: 4 simple predictors, no interactions
- **Model 3**: 4 simple predictors, no interactions, 2 regimes
- **Model 4**: 4 simple predictors, no interactions, 2 regimes, spatial and temporal correlations

Mean Absolute Error of several methods

	Horizon [minutes]	*+10*	*+20*	*+30*	*+40*	*+50*	*+60*	*Total*
Linear regression models	*LR model 1 (two baselines)*	6.096	7.634	8.755	9.821	10.605	11.136	9.008
	LR model 2 (four baselines)	5.961	7.547	8.724	9.807	10.625	11.154	8.969
Baseline predictors incorporated into CCRF models	*Random walk*	6.130	7.667	8.789	10.033	11.072	11.947	9.273
	Historical median	13.090	13.183	13.153	12.994	12.737	12.419	12.931
	Current speed of upstream sensor	7.568	8.746	9.789	10.980	11.995	12.849	10.321
	Current speed of downstream sensor	7.311	8.627	9.543	10.582	11.505	12.277	9.974
CCRF models, in increasing level of complexity	*CCRF model 1 (two baselines)*	6.198	7.703	8.752	9.778	10.597	11.239	9.004
	CCRF model 2 (four baselines)	5.929	7.325	8.333	9.384	10.290	11.061	8.720
	CCRF model 3 (regime-switching)	5.922	7.327	8.329	9.363	10.214	10.891	8.675
	CCRF model 4 (with correlations)	**5.920**	**7.308**	**8.314**	**9.352**	**10.213**	**10.905**	**8.669**

Borrowing strength (missing data issue)

Inherent problem in traffic forecast is missing data (for instance, due to sensor malfunction), and some techniques can not deal with this problem adequately. However, due to the flexibility of CCRF model, missing data problem is mitigated by automatically excluding the corrupted data from the model and using only the existing data to predict. However, since less data are used, we are less certain about the final predictions.

Data preprocessing

The raw data provides us with information about sections' traffic

- **volume** – how many cars passed the sensor during a 30-second interval
- **occupancy** – how long the sensor was occupied during a 30-seconds interval

The data have been aggregated to 5-min intervals, and the speeds were estimated using the following method:

1. We first estimated the average vehicle length by considering only free-flow data and assuming constant speed of 60mph, and further assumed that this length remains constant throughout the period

$$length = 60mph \times median\Big(\frac{occupancy}{volume}\Big)$$

2. We then estimated travel speed using the following function

$$speed = length \times \frac{volume}{occupancy}$$

Conclusions

CCRF is an extremely powerful and flexible model. It allows us to easily:

- Incorporate multiple traffic predictors and different sources
- Exploit spatial and temporal interactions inherent in traffic data
- Model different traffic regimes (free-flow / congested regime, special events such as sports events or accidents, …)
- Obtain simultaneous predictions of large number variables
- Obtain confidence interval for every prediction, providing additional helpful information for travelers
- Extract information about how good certain predictor is, and how strong spatial-temporal correlations in the traffic data are
- Output predictions even when some data are missing (this robustness comes with the price of increased prediction uncertainty)

This novel method can also be applied to other traffic forecasting problems going beyond travel speed predictions.

Acknowledgements

This work was funded in part by the National Science Foundation Grant IIS-0546155. The authors are grateful to Dr. Taek Kwon from University of Minnesota, Duluth, for providing an access to Mn/DOT TMC data.