Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference

A Probabilistic Spatial-Temporal Model and Its Application to Wind Prediction

Nazira Guerrero-Jezzini
Tecnológico de Monterrey,
Ciudad de México (Mexico),
research@naziragj.me

Pablo Ibargüengoytia
Instituto Nacional de Electricidad y Energías Limpias,
Cuernavaca, Morelos (Mexico),
pibar@iie.org.mx

L. Enrique Sucar
Instituto Nacional de Astrofísica, Óptica y Electrónica,
Tonantzintla, Puebla (Mexico), esucar@inaoep.mx

Abstract

Several problems requiere the combination of temporal and spatial reasoning under uncertainty, such as wind prediction for electricity generation in wind farms. In this work we propose a probabilistic spatial-temporal model (PSTM) focused on prediction problems, based on two common properties of these scenarios: sparsity and multivariable mutual information. The proposed spatial-temporal model is essentially a Bayesian network that represents the dependencies between a target variable of interest and a subset of predictor variables in different times and spaces. We developed an algorithm for learning the structure of the model based on a stochastic search of the optimal subset of predictor variables. The proposed model has been applied for wind prediction at different locations in Mexico, using information from several locations at different times. The PSTM is evaluated in terms of predictive accuracy for different time horizons – 1 to 24 hours; and compared to a dynamic Bayesian network (DBN) developed for wind prediction. The performance of the PSTM is in general competitive, and in most cases superior to the DBN.

Introduction

Several problems requiere the combination of temporal and spatial reasoning under uncertainty. For instance, wind prediction in the context of electricity generation in wind farms. In this case, information on climate variables at different times need to be integrated for predicting wind velocity and direction in certain location. As the prediction horizon increases, climate data from other locations needs to be considered. Thus, a spatial-temporal model is required for long term wind forecasting.

Under the framework of probabilistic graphical models (Sucar 2015), there are several techniques for modeling temporal processes such as hidden Markov models and dynamic Bayesian networks, and also for spatial modeling with Markov random fields; but there is no much work that combines both dimensions. In this work we propose a probabilistic spatial-temporal model focused on prediction problems. This model is based on two properties that apply for spatial-temporal prediction problems:

Sparsity: the variable of interest (i.e. wind in an specific location) tends to depend on a few variables across the temporal and spatial dimensions. In previous work (Sanandaji et al. 2015) it has been shown that the wind prediction problem has in general low dimensionality, that is few variables have the largest impact on wind velocity in certain region, so it is possible to achieve good predictions with a low number of variables.

Multivariable mutual information: to select the subset of relevant variables for predicting the variable of interest (V), in general it is better to consider the mutual information between a subset of N variables, that for each variable individually (Kraskov et al. 2005). For instance, it could be that the *influence* of X or Y over V is low, but the combined influence of (X, Y) over V is $high$.

The proposed probabilistic spatial-temporal model (PSTM) is essentially a special type of Bayesian network that represents the dependencies between the variable of interest, V, and a subset of m variables at different times and spaces, $X_1, X_2, ...X_m$. This subset of variables is selected to optimize the predictive accuracy of the model for certain time horizon, T, based on estimating the mutual information between V and $X_1, X_2, ...X_m$. As the number of subsets increases exponentially with respect to the number of variables, m, we developed an stochastic search algorithm for learning the structure of the model; once the *quasi-optimal* structure is determined, the parameters are estimated from data via maximum likelihood. V is then predicted based on the selected spatial-temporal subset using probabilistic inference.

We have applied the proposed model for wind prediction at different locations in Mexico, using information from the same and other locations and at different times. The model is evaluated in terms of predictive accuracy for different time horizons, from 1 to 24 hours; and compared to a dynamic Bayesian network (DBN) model developed for wind prediction (Ibargüengoytia et al. 2014). The results show good predictive performance of the PSTM for most locations and time horizons, and in most cases it is superior to the DBN.

Related Work

As we mentioned in the introduction, to accurately represent a complex system in space-time, a model must exhibit two important characteristics: (i) the representation of the cross-correlation between the spatial and temporal dynamics, and (ii) the properties of correlation among variables through the corresponding dynamic space. To take into account the dynamic correlation through space and time, several models have been proposed (Singh et al. 2010). Some works use non-stationary Gaussian process models to take into account variables' correlation properties in the input space (Garg, Singh, and Ramos 2012; Plagemann, Kersting, and Burgard. 2008). These works are not focused on prediction problems.

There is some previous work that considers spatial and temporal information for wind speed prediction, in particular for short term forecasts. (Larson and Westrick 2006) considered the case of a potential site for a wind farm located in the Columbia River, where weather observations are available from a close site. Another example is the work of (Damousis et al. 2004), for an area with constant thermal winds. (Gneiting et al. 2006) have proposed various models of regime change which represent two dominant wind directions for prediction of wind speed up to 2 hours, with an interesting extension to a probabilistic forecast. More advanced models may be necessary for longer term predictions, as discussed by (Hering and Genton 2010), potentially requiring considerable expertise for the identification of its structure and its parameters. (Gillard and Allard 2013) is based on characteristics of spatial-temporal energy from the wind and use two strategies for obtaining the main direction and speed of propagation. (Gillard and Allard 2013) analyze the structure of stochastic spatial-temporal wind energy for a prediction horizon of up 4 hours.

Previous work on spatial-temporal models in general does not incorporate knowledge about the dependency structure of the domain, and does not represent explicitly the uncertainty on the predictions. This is important for wind prediction, as the decision makers do not only require a point prediction, but also an estimate about the uncertainty (i.e. variance) of this estimate.

A Probabilistic Spatial-Temporal Model

With the aim of developing a probabilistic graphical model that represents relations between time and space, we proposed two main concepts:

Spatial-Temporal Node (STN). A spatial-temporal node, STN, is random variable defined in certain space S and time T. X is a vector containing the record of the events over time, $X = [X_1, X_2, ...X_n]$, and T is the vector containing the time in which each event occurred.

Spatial-Temporal System (STS). A spatial-temporal system is composed of a set of events in discrete space and time, where events at different geographical spaces occur over time. It is represented as a directed acyclic graph, $G = (V, E)$, in which vertices (V) are $STNs$ and the edges (E) define spatial-temporal relations.

We represent a probabilistic spatial-temporal model (PSTM) as a Bayesian network in which each node corresponds to spatial-temporal node at certain space S and time T, X_T^S; and each arc corresponds to a direct probabilistic dependence between X_{T1}^{S1} and Y_{T2}^{S2}. Given that our interest is on prediction, we focus on an specific type of $PSTM$ which represents the dependencies between the variable of interest, at certain space and time, and a set of predictor variables, at different spaces and times. The aim is to reduce the prediction error by considering variables from different spaces and time delays, compared with the traditional models which use a fixed delay and usually an specific space.

Learning the Model

Structure learning. Given that the objective is to build a $PSTM$ to predict an specific variable at certain space and time, the learning algorithm is designed to optimize the prediction accuracy. That is, we want to find a subset of m variables $[X1_{T1}^{S1}, X2_{T2}^{S2}, ...Xm_{Tm}^{Sm}]$ that can best predict the variable of interest, V_T^S. Dependence among variables of different spaces and times is considered to define the set of variables that relate to the objective one. The prediction subset is based on the concept of mutual information, assuming that a subset of variables that present a high mutual information with the target variable will provide good predictions.

So for predicting a target variable V_T^S at time T in space S and prediction horizon H, we want to determine the subset of variables in all spaces, $S_1, S_2, ...S_n$, and for times from $T - H$ to $T - Max$, where $Max > H$ is a predefined maximum temporal range considered for prediction. To select the *optimum subset* of predictor variables we use the criterion of mutual information (MI) between the target variable, Y and a subset of variables (at different times and spaces), $\mathbf{X} = X_1, X_2, ...X_m$, as defined by (Kraskov et al. 2005):

$$MI(Y, \mathbf{X}) = \sum_{Y} \sum_{\mathbf{X}} P(Y, \mathbf{X}) log_2 \frac{P(Y, \mathbf{X})}{P(Y)P(\mathbf{X})} \quad (1)$$

Although the mutual information has been used before for measuring association between variables, and in particular for feature selection (Battiti 1994), in general it is simplified by calculating the individual mutual information between the variable of interest (class) and each of other variables (features). In contrast, this approach is based on the computation of the joint mutual information between the variable of interest and the set of predicting variables, $MI(Y, \mathbf{X})$

As the MI tends to increase with the number of variables, we divided the structure learning algorithm in two phase:

Optimal subset: The optimum subset of predictor variables across space and time is determined by calculating $MI(Y, \mathbf{X})$, for $|X| = L$, where L is the subset size.

Optimal size: The optimal size of the subset is obtained by comparing the optimal subsets of different sizes, based on prediction accuracy[1].

[1] Given the sparsity property we expect that size of the subset will be *small*.

To determine the optimum subset we considered several search strategies, given that the search space increases exponentially with the size of the subsets: Exhaustive search, Hill climbing, Iterative conditional modes and Simulated annealing.

The structure learning process, starts by considering an initial subset size (l), selects the optimum subset for this size and then increases the subset size by one ($i+1$). It determines the optimum subset for $i+1$, and compares it to i in terms of predictive accuracy (or other appropriate measure). If the accuracy improves it continues, otherwise it stops.

Parameter learning. Once the structure is determined, the required parameters are estimated from data. In the case of a prediction $PSTM$ the required parameters are the prior probabilities for each variable in the predictor subset, $P(X_i)$, and the conditional probability of the target variable given the prediction variables, $P(Y \mid X_1, X_2, ...X_m)$. For the prior probabilities we assume a uniform distribution (as in general we will know the value of all these variables, the priors are not relevant). The conditional probability table (CPT), $P(Y \mid X_1, X_2, ...X_m)$, is estimated from data using maximum likelihood.

Prediction

Once the $PSTM$ has been learned, we use it to predict the target variable; selecting the most likely value of the target variable as the one whose probability is the maximum given values of the *optimum* subset of predictor variables:

$$Y_t^* = argmax_{Y_t} P(Y_t^S \mid Y_{t-H}, X1_{t-\delta 1}^{S_1}, ...Xm_{t-\delta m}^{S_m}) \quad (2)$$

where Y_t is the target variable at space S, and $\mathbf{X} = X1, X2, ...Xm$ is the predictor subset determined according to the structure learning algorithm. Each variable in this subset is in space S_i and at time $t - \delta i$ with respect to the prediction time t. Note that $\delta i \geq H$, where H is the prediction horizon. Given that we are discretizing the variables, to improve the accuracy of the predictions we consider changes instead of absolute values. Given a base value, Y_t, we predict the change with respect to this value $\Delta Y = Y_{t+H} - Y_t$, where H is the prediction horizon.

Test Scenario

In order to validate the proposed spatial-temporal model we study the behavior of the wind speed at different geographic locations in Mexico, to determine spatial-temporal relations and forecast the wind speed at each location.

The national meteorological system in Mexico has stations carrying out measurements of meteorological variables in several location in the country. Samples are taken every minute and registered for a 10 minute period average. For the present study we considered three geographical systems (GS) with different number of stations and variables:

GS-12: Twelve stations (located in center-south Mexico), including only *wind speed* at each location.

GS-21: 21 stations and the following variables: *wind speed, wind direction, temperature, relative humidity*.

System	Structure	Search algorithm	learning time (secs)	MAE (km/h) hz=1 hour	hz=2 hours	hz=3 hours	hz=4 hours	hz=6 hours	hz=12 hours	hz=24 hours
GS12	3S	ExhSrch	1.56x10^4	**1.04**	**1.97**	2.89	3.73	5.15	7.52	1.32
		HC	34.82	1.05	1.97	2.9	3.74	5.14	7.52	1.33
		ICM	36.69	1.05	1.98	2.9	3.74	5.14	7.51	1.32
		SA	37.24	1.05	1.98	2.9	3.74	5.14	7.51	1.32
	4S	HC	35.84	1.54	2.28	**2.75**	3.17	3.44	3.85	0.48
		ICM	47.61	1.54	2.29	2.78	3.17	3.43	3.76	0.48
		SA	48.97	1.54	2.29	2.78	**3.17**	**3.43**	**3.76**	**0.48**
GS21	3S	ICM	58.19	2.67	3.53	4.22	5.05	6.77	6.15	3.24
		SA	59.28	2.67	3.53	4.22	5.05	6.77	6.15	3.24
	4S	ICM	72.52	2.56	3.39	4	4.78	6.53	6.89	2.65
		SA	73.3	**2.56**	**3.39**	**4**	**4.78**	**6.53**	**6.89**	**2.65**
GS35	3S	ICM	94.93	3.26	4.06	4.85	5.98	5.91	4.82	2.49
		SA	95.35	3.26	4.06	4.85	**5.98**	**5.91**	4.82	2.49
	4S	ICM	111.1	3.17	3.96	4.78	6.06	7.42	4.15	0.34
		SA	119.8	**3.17**	**3.96**	**4.78**	6.06	7.42	**4.15**	**0.34**

Figure 1: Summary of the prediction error in terms of MAE.

GS-35: 35 stations with the following variables per station: *wind speed, wind direction, temperature, relative humidity, atmospheric pressure, radiation*.

The first scenario is used to compare the different heuristic search approaches with the exhaustive search, and to contrast our model with a DBN. The other two, more complex scenarios, evaluate the capacity to deal with a larger set of spaces and variables. For training and evaluation we included the data of the different stations, averaged per hour. For each scenario we used data for a number of months for training, and data from other different months for testing.

Common metrics for prediction were used: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE) and Error of the Instrument (EI).

Experiments and Results

For predicting the wind velocity in each location of the three scenarios, we considered all the variables in all the locations as predictor variables, within a time range of $[t - H, t - 36]$, where H is the prediction horizon in hours. We experimented for $H = 1, 2, 3, 4, 6, 12, 24$. A model was learned for each location and time horizon, so in total $12 \times 7 = 84$ models were learned for GS-12, $21 \times 7 = 147$ for GS-21, and $35 \times 7 = 245$ for GS-35.

In these experiments we considered a fixed size for the prediction subsets, with two (3S) and three variable (4S) –three and four considering the target variable–, learning and evaluating all the models for these cases –a total of $(84 + 147 + 245) \times 2 = 958$ models. We leave the second phase of the structure learning process –establishing the optimal size of the subset– as future work.

The results are summarized in terms of Mean Absolute Error in Figure 1 for the three scenarios (GS-12, GS-21 and GS-35), two structure sizes (there and four variables), seven prediction horizons (hz) and the different search strategies (the other metrics have a similar behavior and are not included due to space limitations). The table also includes the average learning time for each search strategy in each scenario.

The time to learn 12 models with exhaustive search is in

		Time horizon = 4 hours							
		DBN				PSTM			
Ey	EMA	MAE (km/h)	RMSE (%)	NRMSE (%)	Inst error (%)	MAE (km/)	RMSE (%)	NRMSE (%)	Inst error (%)
1	E01 Acapulco	4.01	5.36	1.26	22.93	2.51	4.05	1.83	5.9
2	E02 Cd. Altamirano	2.25	2.77	1.20	17.85	2.06	6.13	3.81	16.71
3	E03 Petacalco	3.73	4.76	1.43	19.02	2.93	4.38	1.94	6.37
4	E04 IMTA	4.02	4.73	1.68	23.66	3.3	4.55	2.05	9.13
5	E05 Tepoztlan	2.40	2.98	1.18	16.13	2.35	5.57	1.93	7.31
6	E06 Nochixtlan	5.92	7.54	1.34	20.01	4.98	6.73	1.94	10.39
7	E07 Pinotepa Nacional	3.21	3.87	1.17	19.45	2.48	6.5	2.87	12.89
8	E08 Puerto Angel	5.39	6.34	1.92	24.30	2.06	2.92	0.93	1.46
9	E09 Izucar	5.96	7.60	1.37	16.83	4.45	5.71	1.54	7.31
10	E10 UT Tecamachalco	4.88	7.77	1.11	13.04	4.34	5.75	1.81	7.24
11	E11 Cd. Aleman	3.78	4.66	1.54	18.24	2.05	2.93	1.73	7.03
12	E12 Citlaltepec	0.95	1.23	1.34	8.60	2.46	3.4	2.04	7.66

Figure 2: Comparison of the results obtained with the proposed approach (PSTM) against a DBN with prediction horizon of 4 hours, for the 12 locations in GS-12 scenario.

the order of $15,600$ secs. for subsets of 3 variables (AMD 2 GHz, RAM 12 GBytes). However, with the other search strategies is in the order of 30 seconds to two minutes for all the scenarios. Note that the increase in learning time is not significant with the larger scenarios (GS-21, GS-35), so with these strategies it is feasible to learn the models even for more complex problems.

If we analyze the resulting structure of the learned spatial-temporal models is interesting to observe that: (i) there is a wide variety of space and time for the variables on the predictor subset, (ii) although there is some tendency with respect to the spaces selected for predicting each location, these vary according to the prediction horizon. In general the learned structures incorporate variables in close regions of the region of interest, which are consistent with the dominant winds. An interesting example is station 8 (Oaxaca), which is influenced by station 12 in Veracruz to the north. Oaxaca is a region with strong winds that usually come from the north.

Comparison against DBNs

We compared the predictions of the proposed probabilistic spatial-temporal model (PSTM) against a dynamic Bayesian network (DBN) which has been developed for wind prediction (Ibargüengoytia et al. 2014). A DBN model was learned for the GS-12 scenario considering the 12 wind velocity variables using the same data set as for the PSTM for training and testing. See (Ibargüengoytia et al. 2014) for more details .

The prediction was performed for time horizons of one to four hours, and evaluated with the same metrics. We present a comparison for $H = 4$, depicted in Figure 2. We can appreciate that for most locations superior results are obtained with the $PSTM$ for all metrics; and for some locations the difference is significant. The same pattern repeats for the other time horizons that were evaluated.

Conclusions and Future Work

We developed a probabilistic spatial-temporal model for prediction problems, that encodes the dependencies between a target variable and a subset of predictor variables at different times and spaces. We have applied the proposed model for wind prediction at different locations in Mexico. The results are competitive for horizons of up to 24 hours, and superior to a DBN model. The assumption of sparsity could be seen as a limitation of the proposed approach, as the number of parameters grows exponentially with the number of predictor variables. An alternative is to consider the use of canonical models (such as the Noisy-OR and Noisy-AND), where the number of parameters grows linearly.

As future work we will evaluate the second phase of the structure learning algorithm to determine the optimal number of variables in the model.

References

Battiti, R. 1994. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks* 5(4):537–550.

Damousis, G.; Alexiadis, M.; Theocharis, J.; and Dokopoulos, P. 2004. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion* 19:–.

Garg, S.; Singh, A.; and Ramos, F. 2012. Learning non-stationary space-time models for environmental monitoring. In *AAAI*.

Gillard, R., and Allard, D. 2013. Spatio-temporal propagation of wind power prediction errors. *Wind Energy* 16:999–1012.

Gneiting, T.; Larson, K.; Westrick, K.; Genton, M.; and Aldrich, E. 2006. Calibrated probabilistic forecasting at the satellite wind energy center: the regimes witching space-time method. *Journal of the American Statistical Association* 101:968–979.

Hering, A., and Genton, M. 2010. Powering up with space-time wind forecasting. *Journal of American Statistical Association* 105:92–104.

Ibargüengoytia, P. H.; Reyes, A.; Romero, I.; Pech, D.; García, U.; Sucar, L.; and Morales, E. F. 2014. Wind power forecasting using dynamic bayesian models. In Gelbokh, A., ed., *Advances in Soft Computing, part II - MICAI 2014, LNAI 8857"*. Berlin Heidelberg: Springer-Verlag.

Kraskov, A.; Stogbauer, H.; Andrzejak, R.; and Grassberger, P. 2005. Hierarchical clustering based on mutual information. *EPL (Europhysics Letters)* 70(2).

Larson, K., and Westrick, K. 2006. Short-term wind forecasting using off-site observations. *Wind Energy* 9:55–62.

Plagemann, C.; Kersting, K.; and Burgard., W. 2008. Non-stationary gaussian process regression using point estimates of local smoothness. In *European Conference on Machine Learning*.

Sanandaji, B.; Tascikaraoglu, A.; Poolla, K.; and Varaiya, P. 2015. Low-dimensional models in spatio-temporal wind speed forecasting. In *American Control Conference*.

Singh, A.; Ramos, F.; Durrant-Whyte, H.; and Kaiser, W. 2010. Modeling and decision making in spatio-temporal processes for environmental surveillance. In *IEEE International Conference on Robotics and Automation*.

Sucar, L. 2015. *Probabilistic Graphical Models: Principles and Applications*. London: Springer-Verlag.