Spatiotemporal Literature Review

Bo Feng
Department of Intelligent Systems Engineering
Indiana University - Bloomington
fengbo@iu.edu

June 6, 2021

Abstract

Spatiotemporal pattern prediction is one of the emerging topics in Deep Learning applications. In this report, we recall some technical background in studies about spatiotemporal pattern prediction and review some essential academic works of literature in related fields. We also categorize research works by their application domains and highlight the applications on addressing scientific problems.

1 Introduction

Spatial and temporal attributes have played an essential role in addressing scientific issues mathematically and statistically with large volumes of data in real problems. Authors of this book [1] summarized some initial efforts by utilizing spatiotemporal features for scientific interpretation and prediction.

Most recently, it is a prevailing method to make predictions by modeling the spatiotemporal dynamics for domain science problems. This popularity is because large volumes of data are increasingly collected in the vast majority of domains including, social science, epidemiology, transportation, and geoscience.

In this writeup, we review some technical background for spatiotemporal research, especially in deep learning models in Section 2. Then, we summary some prior research related to spatial temporal modeling in Section 3.

2 Background Technique Review

2.1 Convolutional Neural Networks

Convolutional operations on time series data can be used to predict patterns of object movement. Karatzoglou*et al.* build models for GPS location signals for tracking human movement patterns [2].

Traffic forecasting is one of the most popular applications in spatial temporal forecasting. The challenges of traffic forecasting are related to varying traffic

patterns and spatial dependencies of traffic networks. Cui et al. proposed a model named Graph Convolutional Long Short-Term Memory Neural Network (TGC-LSTM), which modeled traffic between roadways and forecast traffic state of the road networks [3]. In this paper, the modeling task is to learn a function that can predict the traffic pattern at $T + 1$ time based on a series of history signals from time 1 to T. In this paper, the k-top of the neighborhood and a free-flow reachable are modeled as graph nodes and edges. To apply some node traffic-specific restrictions on nodes' extracted features, a customized loss function is defined with a L1-norm and L2-norm functions.

2.2 Graph Representation with Neural Networks

Compared to traditional convolutional neural networks (CNN), graph-based convolutional networks (GCN) are becoming a powerful tool to model problems. GCN is especially useful for applications where data are formed in non-Euclidean geometry systems and are represented in graphs in which dependencies of nodes are complex. Action recognition is one type of spatial temporal problems in which a sequence of body actions can be modeled in a spatial temporal graph. Yan et al. proposed a Spatial Temporal Graph Convolutional Networks (ST-GCN) to model dynamic skeletons [4]. In this paper, a graph is formed as nodes defined in body joints and edges based on bones or natural connections in human bodies. A spatial temporal graph is a skeleton sequence of body actions. This is a classification problem in which the output of the model is categorized as a type of movement such as running or jumping.

Geo-based service demand prediction is another type of spatial temporal problems. Spatial temporal graph is popular to be used to predict ride-hailing demand. Geng et al. proposed a multi-graph convolutional model for ride-hailing demand forecasting [5]. In this paper, the input of the model is a sequence of ride-demand history maps from $t - (T + 1)$ time to t time as the input $(X^{t-T+1}, X^{t-T+2}, \ldots, X^t)$, and the output of the model consists of a predicted demand map at t+1 time X^{t+1}. The problem is modeled as multigraphs, each of which has its separated correlation weights. The input data is aggregated on maps partitioned into grids of 1km x 1km squared regions.

2.3 Recurrent Neural Networks

Recurrent neural networks (RNN) and its successful model type—Long short-term memory (LSTM) was designed and developed to learn information overtime via recurrent neural networks [6].

2.3.1 Long Short-Term Memory

LSTM has a wide range of applications to capture temporal dependencies. Combining with convolutional neural network layers, this type of neural network is capable of learning patterns for spatial temporal problems. This method is widely used in air pollution forecasting[7] and prediction of complex physical systems [8].

Typically, in this method, the one-dimensional convolution operation is used to extract features of input along time series to capture spatial dependencies, and N layers of LSTM concatenated in the networks are used to capture temporal dependencies.

Convolutional LSTM (ConvLSTM) as a special type of LSTM replaces the conventional matrix multiplication with the convolutional multiplication for 2D input [9]. The ConvLSTM was originally designed to capture spatiotemporal dynamics.

2.3.2 Gated Recurrent Unit

Similar to LSTM, gated recurrent unit (GRU) is another type of RNN, which draw great attention since its development [10]. GRU was firstly proposed by[11], which has a similar structure to the LSTM unit without a separate memory cell. According to [10], it is hard to conclude which type of RNN can perform better in general between LSTM and GRU.

Li et al. proposed a neural network model adopting GRU for car traffic spatial temporal forecasting [12]. In this paper, the learning task is to learn a function that maps from a sequence of T history time signals to a sequence of T time signals. Their solution is a synthetic method consisting of 1) a spatial dependency modeling using random walks on graph and diffusion convolution, and 2) a temporal dependency modeling using GRU.

2.3.3 Hierarchical Recurrent Neural Network

A hierarchical recurrent neural network is a general form of stacking RNN layers, which was developed to capture long term dependency [13]. Stacking LSTM is a particular form of hierarchical recurrent neural network. The hierarchical recurrent neural network can address issues of modeling spatial temporal dynamics. In 2015, Du et al. proposed to use hierarchical RNN for human action recognition [14]. This problem is almost the same as defined in Yan's paper [4]. Compared to Yan's paper, in Du's paper, a human skeleton is decomposed into five parts: two arms, two legs, and one trunk. Each skeleton part is fed into each bidirectional recurrent network (BRNN). To model the spatial dynamics, such as arm-trunk or leg-trunk correlations, they combine the trunk subnet with each of the other four subnet of BRNN. In the implementation, LSTM unit is adopted in the recurrent layer. The final output is categorized as one type of all actions.

2.4 Encoder-Decoder Architecture

2.4.1 Restricted Boltzmann Machines (RBM)

Hinton proposed the RBM model, an energy-based model consisting of only two layers in a network. One layer is called visible node layer, and another is called hidden node layer. This model is trained using the contrastive divergence, which is the difference between two Kullback-Liebler divergences [15].

2.4.2 Autoencoder and Variational autoencoder

Autoencoder is one of the typical models that follow the encoder-decoder architecture, in which an encoder produces latent variables and the decoder takes the output of an encoder as input. According to [16], the autoencoder can learn low dimensional features that are similar to the results of the PCA algorithm. The process can be expressed in a simply math formula: $x = g(f(x))$, in which f is the encode function and g is the decode function.

Variational autoencoder (VAE) extends the idea of the encoding-decoding process with latent variables to using variational distributions. Some previous work presented its applications for images [17, 18]. VAE is basically a Bayes' process of generating data from a prior distribution as follows: $p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$. Compared to the Monte Carlo EM algorithm, VAE is more efficient in handling large datasets.

2.4.3 RNN Autoencoder

Sequence to sequence translation using an RNN-based Encoder-Decoder architecture was proved to be a powerful tool used in linguistic grammar analysis and machine translation in around 2014 [11, 19]. According to Cho et al.'s papers, in a typical RNN encoder-decoder, an encoder is an RNN that reads a sequence of symbols and produces a summary sequence, then the summary sequence is fed into another RNN decoder. An aforementioned paper that predicts traffic[12] also adopts the encoder-decoder architecture with GRU. Google's attention-based transformer mechanisms is also a type of encoder-decoder networks without RNN or CNN [20, 21].

2.5 Parallel Time Steps and Attention Mechanism

2.5.1 WaveNet and Temporal Convolutional Network

One of the major efforts to model time steps during recent years is to enable the model trainable in parallel. The very first representative is the WaveNet [22], which was proposed by a Google Mind team. It is designed to be a generative model for audio and text based on the PixelCNN architecture [23]. WaveNet models the joint probability of $\mathbf{x}$ as a product of conditional probabilities from all previous states: $p(\mathbf{x}) = \prod_{t=1}^{T} p(x_t|x_1, x_2, \ldots, x_{t-1})$. To model temporal dynamics, WaveNet uses dilated convolutions passing information layer by layer.

Temporal convolutional network (TCN) was introduced from a few work, including [24, 25, 26]. Temporal convolutional network inherits the dilated convolutional from conditional WaveNets, in which a conditional distribution is modeled from: $p(\mathbf{x}|h) = \prod_{t=1}^{T} p(x_t|x_1, x_2, \ldots, x_{t-1}, h)$. TCN is an efficient model respected to the time domain due to its parallel model structures and is suitable for regression-type problems.

Table 1: Spatiotemporal application domains

Domains	Related work
Scientific	[7], [30], [9], [31], [32], [33], [34], [35], [36], [37], [38]
Transportation and traffic	[2], [3], [12] [5], [39]
Multimedia and online service	[29] [4], [27], [22], [23], [14]
Natural language processing	[40], [20], [21], [11] [19]

2.5.2 Attention and Transformer model(s)

Attention in neural network models is a mechanism that weights the input states to the output states rather than using a fixed-sized feature vector. Attention mechanism was originally proposed to improve the decoder in an Encoder-Decoder architecture [27]. The original transformer model extends this attention mechanism to encoder-decoder stacks in multiple heads [20]. The successor of Transformer, BERT [21] is a framework for pre-training and tuning the most important parameters from a Transformer network, such as Transformer blocks, hidden sizes, and self-attention heads.

2.6 Generative Adversarial Network (GAN)

A generative adversarial network (GAN) is a new set of generative models [28]. A typical GAN consists of two major neural networks: a generator G and a discriminator D. It simulates a two-player game for G and D, where G tries to fool D, and D tries to tell the truth. GAN is getting popular in research about spatiotemporal topics. Zhu *et al.* [29] proposed a GAN model for detecting spatiotemporal events, in which both the generator and discriminator are built with one LSTM layer.

3 Spatiotemporal Research Work Review

Four subjects are related to this project: 1) predicting locations with machine learning techniques; 2) spatiotemporal modeling in a broad range of applications; 3) advanced dynamic pattern mining and prediction in visual applications; 4) extreme event prediction in other areas of research.

3.1 Convolutional Methods in Predicting Epicenters

Estimating and predicting the epicenters of earthquakes has a long history. Geophysics, Geology, and Seismology have developed various tools and analytical functions to predict epicenters from datasets. In 1997, Bakun and Wentworth suggested using Modified Mercalli intensity datasets for Southern California earthquakes to bound the epicenter regions and magnitudes [33]. In 1998, Pulinets

proposed predicting epicenters of strong earthquakes with the help of satellite-sounding systems. Scientists from Greece had illustrated a successful project which predicted the prominent aspects of earthquakes using seismic electric signals [34]. Recently, Guangmeng *et al.* attempted to predict earthquakes with satellite cloud images and revealed some possibilities of predicting earthquakes using geophysics data [35]. Zakaria *et al.* presented their work of predicting epicenters by monitoring precursors, such as crustal deformation anomalies and thermal anomalies, with remote sensing techniques [36]. These studies either used only too little data or too simple analytical models. In this project, we cover a date range from 1950 to 2019 and transform the dataset into a dense series of images. Rundle *et al.* adopt the same data processing approach from us [37].

3.2 Spatiotemporal Dynamics and Generative Models

Due to large volumes of data and advanced models been developed, spatiotemporal dynamics modeling is increasingly popular in many domains. Cui *et al.* proposed to use graph convolutional long short-term memory neural networks to predict traffic via capturing spatial dynamics from the car traffic patterns [3]. Li *et al.* utilized a seq2seq neural network architecture to capture spatial and temporal dependencies for traffic forecasting by incorporating a diffusion filter in convolutional recurrent layers [12]. FUNNEL was a project proposed by Matsubara [38]. It was designed to use an analytical model and a fitting algorithm for discovering spatial-temporal patterns of epidemiological data.

3.3 Visual Pattern Prediction

Lotter *et al.* presented a model to predict video frames with deep predictive coding networks [41], which was based on the ConvLSTM2D network module with specific top-down states updating algorithm. [32] is another example in predicting video frames. The authors of this work presented the effectiveness of modeling object motion via predicting future object pixels. For example, a ball moves, and a block falls. These models are successful for predicting contiguous and dense image frames. By contrast, the earthquake data are very sparse, and the extreme shocks are rare in terms of probability.

3.4 Extreme rare event prediction

Laptev *et al.* [39] proposed their modeling to predict rare trip demands for ride-hailing service. In that paper, they built an end-to-end architecture using joint modeling by combining LSTM autoencoder and LSTM predictor networks. They showed their forecasting capability on a Uber's public dataset. Geng *et al.* [5] proposed another model to forecast the ride-hailing demand using graph-based recurrent neural networks, in which graphs are defined by road networks with Euclidean and non-Euclidean distances. Zhu *et al.* [29] proposed a GAN based model for detecting streaming spatiotemporal events.

4 Conclusion

In summary, we present some reviews and summaries about the technical background in deep learning for spatiotemporal research. Some applications are widely adopted for video/image and text/speech. However, the deep learning neural networks are also applicable for addressing scientific problems where the inner principles for pattern recognition and prediction are shared across.

References

[1] George Christakos. *Modern Spatiotemporal Geostatistics*. Oxford University Press, November 2000.

[2] Antonios Karatzoglou, Nikolai Schnell, and Michael Beigl. A Convolutional Neural Network Approach for Modeling Semantic Trajectories and Predicting Future Locations. In Věra Kůrková, Yannis Manolopoulos, Barbara Hammer, Lazaros Iliadis, and Ilias Maglogiannis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2018*, Lecture Notes in Computer Science, pages 61–72, Cham, 2018. Springer International Publishing.

[3] Zhiyong Cui, Kristian Henrickson, Ruimin Ke, and Yinhai Wang. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[4] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3656–3663, 2019.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[7] Chiou-Jye Huang and Ping-Huan Kuo. A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities. *Sensors*, 18(7):2220, 2018.

[8] Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature*, 568(7753):526–531, 2019.

[9] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 802–810. Curran Associates, Inc., 2015.

[10] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555 [cs]*, December 2014.

[11] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078 [cs, stat]*, September 2014.

[12] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *arXiv:1707.01926 [cs, stat]*, July 2017.

[13] Salah El Hihi and Yoshua Bengio. Hierarchical recurrent neural networks for long-term dependencies. In *Advances in Neural Information Processing Systems*, pages 493–499, 1996.

[14] Yong Du, Wei Wang, and Liang Wang. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1110–1118, 2015.

[15] Geoffrey E. Hinton. A Practical Guide to Training Restricted Boltzmann Machines. In Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, volume 7700, pages 599–619. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[16] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808. Citeseer, 2007.

[17] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv:1312.6114 [cs, stat]*, December 2013.

[18] Tim Salimans, Diederik Kingma, and Max Welling. Markov Chain Monte Carlo and Variational Inference: Bridging the Gap. In *International Conference on Machine Learning*, pages 1218–1226. PMLR, June 2015.

[19] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to Sequence Learning with Neural Networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3104–3112. Curran Associates, Inc., 2014.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, \Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019.

[22] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*, September 2016.

[23] Aaron Van Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel Recurrent Neural Networks. In *International Conference on Machine Learning*, pages 1747–1756. PMLR, June 2016.

[24] Colin Lea, René Vidal, Austin Reiter, and Gregory D. Hager. Temporal Convolutional Networks: A Unified Approach to Action Segmentation. In Gang Hua and Hervé Jégou, editors, *Computer Vision – ECCV 2016 Workshops*, Lecture Notes in Computer Science, pages 47–54, Cham, 2016. Springer International Publishing.

[25] Colin Lea, Michael D. Flynn, Rene Vidal, Austin Reiter, and Gregory D. Hager. Temporal Convolutional Networks for Action Segmentation and Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165, 2017.

[26] Anastasia Borovykh, Sander Bohte, and Cornelis W. Oosterlee. Conditional Time Series Forecasting with Convolutional Neural Networks. *arXiv:1703.04691 [stat]*, September 2018.

[27] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv:1409.0473 [cs, stat]*, May 2016.

[28] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[29] Shixiang Zhu, Henry Shaowu Yuchi, and Yao Xie. Adversarial Anomaly Detection for Marked Spatio-Temporal Streaming Data. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8921–8925, May 2020.

[30] Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature*, 568(7753):526–531, April 2019.

[31] Xiong Zhang, Jie Zhang, Congcong Yuan, Sen Liu, Zhibo Chen, and Weiping Li. Locating earthquakes with a network of seismic stations via a deep learning method. *Scientific Reports*, 10(1):1941, December 2020.

[32] Chelsea Finn, Ian Goodfellow, and Sergey Levine. Unsupervised Learning for Physical Interaction through Video Prediction. *arXiv:1605.07157 [cs]*, October 2016.

[33] W. H. Bakun and C. M. Wentworth. Estimating earthquake location and magnitude from seismic intensity data. *Bulletin of the Seismological Society of America*, 87(6):1502–1521, December 1997.

[34] S. A Pulinets. Strong earthquake prediction possibility with the help of topside sounding from satellites. *Advances in Space Research*, 21(3):455–458, January 1998.

[35] G. Guangmeng and Y. Jie. Three attempts of earthquake prediction with satellite cloud images. *Natural Hazards and Earth System Sciences*, 13(1):91–95, January 2013.

[36] Zahra Alizadeh Zakaria and Farshid Farnood Ahmadi. Possibility of an earthquake prediction based on monitoring crustal deformation anomalies and thermal anomalies at the epicenter of earthquakes with oblique thrust faulting. *Acta Geophysica*, 68(1):51–73, February 2020.

[37] John B. Rundle, Andrea Donnellan, Geoffrey Fox, James P. Crutchfield, and Robert A. Granat. Nowcasting Earthquakes: Imaging the Earthquake Cycle in California with Machine Learning. http://www.essoar.org/doi/10.1002/essoar.10506614.1, March 2021.

[38] Yasuko Matsubara, Yasushi Sakurai, Willem G. van Panhuis, and Christos Faloutsos. FUNNEL: Automatic mining of spatially coevolving epidemics. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 105–114, New York, NY, USA, August 2014. Association for Computing Machinery.

[39] Nikolay Laptev, Jason Yosinski, Li Erran Li, and Slawek Smyl. Time-series Extreme Event Forecasting with Neural Networks at Uber. In *International Conference on Machine Learning*, page 5, 2017.

[40] Fan Chung and Olivia Simpson. Computing heat kernel pagerank and a local clustering algorithm. *European Journal of Combinatorics*, 68:96–119, February 2018.

[41] William Lotter, Gabriel Kreiman, and David Cox. Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. In *International Conference of Learning Representations (ICLR)*, February 2017.