Machine Learning Meets Big Spatial Data (Revised)

Ibrahim Sabek
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, USA
sabek@mit.edu

Mohamed F. Mokbel
Department of Computer Science and Engineering
University of Minnesota, USA
mokbel@umn.edu

***Abstract*—The proliferation in amounts of generated data has propelled the rise of scalable machine learning solutions to efficiently analyze and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, e.g., GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, drug addiction, among others. Consequently, major research efforts are exerted to support efficient analysis and intelligence inside these applications by either providing spatial extensions to existing machine learning solutions or building new solutions from scratch. In this 90-minutes seminar, we comprehensively review the state-of-the-art work in the intersection of machine learning and big spatial data. We cover existing research efforts and challenges in three major areas of machine learning, namely, data analysis, deep learning and statistical inference. We also discuss the existing end-to-end systems, and highlight open problems and challenges for future research in this area.**

I. Introduction

There has been a recent wide deployment of machine learning (ML) solutions, with their different areas (e.g., data analysis, deep learning), in various big data applications, including public health [24], information extraction [67], data cleaning [53], among others. Meanwhile, spatial applications have witnessed unprecedented explosion in the amounts of generated and collected data. For example, medical devices produce spatial images (X-rays) at a rate of 50 PB per year, while a NASA archive of satellite earth images has more than 500 TB. To efficiently process such tremendous amounts of spatial data, researchers and developers worldwide have proposed either spatial extensions to existing machine learning systems (e.g., Azure Geo AI [3]) or new end-to-end solutions (e.g., ESRI ArcGIS [14]). Such extensions and new solutions have motivated a wide variety of applications in biology [71], environmental science [72], climatology [17], among others.

Scope. In this seminar, we aim to provide a comprehensive review of existing machine learning systems and approaches that efficiently support big spatial data. Figure 1 depicts the landscape of the intersection between machine learning and big spatial data worlds that will be covered in this seminar. The horizontal axis in Figure 1 represents the type of each machine learning solution, whether it takes the distinguishing spatial data properties into account or not, while the vertical axis represents the type of application employing such machine

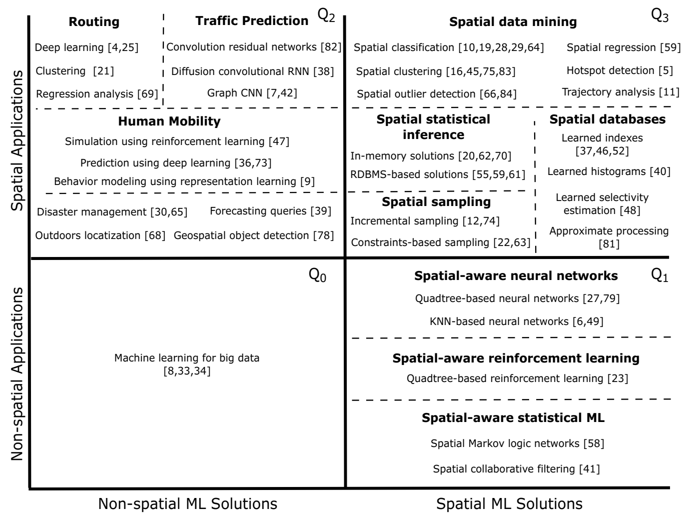

Fig. 1. Landscape of Machine Learning for Big Spatial Data.

learning solution, whether the application is spatial or not. We mainly focus on the three quarters Q_1, Q_2, and Q_3 in Figure 1 because they cover the spatial dimension in the machine learning solutions and/or the big data applications. We skip the quarter Q_0 as it is already covered by previous SIGMOD tutorials about the techniques and challenges in machine learning for big data in general [8], [33], [34].

Related tutorials. There were two previous tutorials [1], [13] related to this seminar. The first tutorial [13] focused on big spatial data management. However, unlike this tutorial, our seminar aims to combine the two worlds of *scalable machine learning* and *big spatial data* together, which is beyond just applying techniques from one area to another. The second tutorial [1] focused only on learned spatial indexes, which as shown in our seminar, is only one category in the quarter Q_3 of our proposed landscape.

Prior offerings. The authors have presented a 90-minutes tutorial about the same topic in the VLDB 2019 and ICDE 2020 conferences [56], [57]. However, this seminar provides a thorough revision to the previously presented landscape and adds more recent works in the three quarters Q_1, Q_2, and Q_3.

II. Seminar Outline

Figure 2 gives the **90-minutes** seminar outline, composed of five parts. The first part motivates the need for machine learning systems to support big spatial data, and provides the

[1]This work is partially supported by the National Science Foundation, USA, under Grants IIS-1907855, IIS-1525953, CNS-1512877 and CCF-2030859.

basic background on these two worlds (Section II-A). The second, third, and fourth parts delve into the ongoing machine learning efforts and challenges in the quarters Q_1, Q_2, and Q_3 from Figure 1, respectively (Sections II-B to II-D). In each of these three quarters, we explain the main ideas, architectures, strengths and weaknesses of existing machine learning solutions. We also highlight the strong bond between spatial data management and spatial machine learning workflows, discuss the related technical challenges, and outline the open research opportunities. The fifth part reviews the existing end-to-end systems for big spatial data analysis (Section II-E).

A. Part 1: Spatial Data and ML Synergy

This part advocates for the need to develop machine learning systems and techniques for big spatial data that go beyond simple extensions of existing work for general data. We start by describing some motivating applications, introducing the world of big spatial data, and discussing its machine learning related concepts. We then quickly review the landscape of spatial machine learning systems, algorithms, applications, and needs, which will be heavily discussed in the next parts.

B. Part 2: Spatial ML Solutions for Non-spatial Apps

This part covers the role of injecting the spatial awareness inside the underlying machine learning algorithms used in non-spatial applications (e.g., knowledge base construction [58], recommendation systems [41], computer vision [27]) to improve the performance of these applications. We start by highlighting how the spatial data management techniques improve the performance of various tasks in neural networks and reinforcement learning when applying on big spatial data. For example, Quad-tree partitioning [18] is used for: (a) balancing the convolution computation in Convolutional Neural Networks (CNN) for object detection applications [27], (b) efficient automatic features extraction and matrix factorization operations inside deep learning models [79] and (c) parallelizing the reinforcement learning computation for motion planning [23]. Meanwhile, k-nearest neighbor operations are used to efficiently build specific neural network architectures from big spatial datasets [6], [49]. Then, we discuss the improved spatial variations of other statistical machine learning techniques (i.e., not deep learning) used inside knowledge base construction [58], [60] and recommendation [41] models, while assuring their impact in obtaining more accurate outputs.

C. Part 3: Non-spatial ML Solutions for Spatial Apps

This part covers the usage of existing machine learning techniques, without spatial variations, as "black boxes" in improving the performance of spatial applications. We start by discussing the recent machine learning techniques used inside three specific core applications; routing, traffic prediction and human mobility. For routing, we show the deep learning [25] and regression analysis [69] techniques used to prepare the routing meta-data (e.g., finding weights of routes). We also present the incremental learning [4] and clustering [21] approaches that are used to make routing maps and perform the routing itself, respectively. For traffic prediction, we present examples of its existing deep learning [7], [38], [42], [82], as well as reinforcement learning [76] approaches in details. For human mobility, we discuss its simulation using reinforcement learning [47], prediction using federated [36] and deep learning [73], and behavior modeling using representation learning [9]. Finally, we give a brief about the machine learning approaches used in other spatial applications including disaster managment [30], [65], outdoors localization [68], forecasting queries [39], and geospatial object detection [78].

- **Part 1: Spatial Data and ML Synergy (10 mins)**
  - Importance of ML with big spatial data
  - Quick review of spatial ML landscape
- **Part 2: Spatial ML Solutions for Non-spatial Apps (20 mins)**
  - Spatial-aware neural networks and reinforcement learning
  - Spatial-aware statistical ML models (not deep learning)
- **Part 3: Non-spatial ML Solutions for Spatial Apps (25 mins)**
  - ML for routing, traffic prediction and human mobility (In-depth)
  - ML for disaster analysis, localization and object detection (Brief)
- **Part 4: Spatial ML Solutions for Spatial Apps (25 mins)**
  - Learned spatial data management operations
  - Scalable spatial data mining techniques
  - Scalable spatial inference and sampling techniques
- **Part 5: End-to-end Spatial Data Analysis Systems (10 mins)**
  - Spatial support in existing big data analysis systems
  - Full-fledged big spatial data analysis systems

Fig. 2. Seminar Outline (90 minutes)

D. Part 4: Spatial ML Solutions for Spatial Apps

This part covers the research efforts of learned spatial data management operations and scalable spatial data analysis techniques. For spatial data management, we cover recent works in learning spatial indexes [1], [37], [46], [52], multi-dimensional histograms [40], selectivity estimation [48] and approximate processing [81]. For spatial data analysis, we touch on the efforts for scaling up the performance of three main categories: (1) *Spatial data mining*: common operations in this category include spatial outlier detection [66], [84], spatial classification [10], [19], [28], [29], [64], spatial regression [59], spatial clustering [16], [45], [75], [83], hotspot detection [5], and trajectory analysis [11]. (2) *Spatial statistical inference*: existing spatial inference approaches are categorized into: (a) *in-memory* solutions, where the input dataset of the inference model is first spatially partitioned into a grid. Then, each partition is analyzed using a Bayesian spatial process model (e.g., [20]). Finally, an approximate posterior inference for the entire dataset is obtained by optimally combining the individual posterior distributions from each partition [20], [62], [70]. (b) *RDBMS-based* solutions, where the assumption of fitting the whole model data in memory is no longer valid. Hence, RDBMSs are exploited to support scalable spatial inference computation (e.g., TurboReg [59] and Flash [55], [61]). (3) *Spatial sampling*: existing sampling techniques over big spatial data can be either incremental (i.e., samples are refined over many iterations) [12], [74] or satisfying certain locality constraints (e.g., zooming level) [22], [63].

E. Part 5: End-to-end Spatial Data Analysis Systems

This part covers the big spatial data analysis systems from two aspects: (1) The research efforts of adding spatial support in existing big data analysis systems, which are either: (a) in the form of add-ons libraries and tools that enable processing spatial data with classical operations (e.g., clustering, classification). Examples include spatial extensions to Spark core (e.g., Simba [77], Magellan [43], GeoSpark [80], GeoMesa [26], UlTraMan [11]) to enable using Spark MLib [44] with spatial data, ESRI spatial data analysis extensions for Hive [15], and PostGIS [50] that can be used along with MADLib [24] to support spatial analytics for PostgreSQL [51], or (b) in the form of built-in native support of spatial analysis operations (e.g., hot spot detection, spatial co-location) inside existing data analysis engines. (2) The research efforts of providing full-fledged big spatial data analysis systems and tools. In such systems, all execution steps in any data analysis operation are optimized for efficient and scalable processing of spatial data. We will classify existing work based on the underlying architecture, which could be either (a) *in-memory systems* (e.g., CrimeStat [35], GeoDa [2], PySAL [54]), (b) *RDBMS-based systems* (e.g., ESRI ArcGIS [14], Flash [61]), or (c) *cloud-based services* (e.g., IBM PAIRS [31]).

III. Target Audience and Relevance to MDM

This seminar targets researchers, developers, and practitioners, who are interested in the intersection area between large-scale machine learning and big spatial data. Research in this area recently becomes very active in the database and spatial communities in general, and in the MDM community in particular. Many of the research efforts covered in this seminar were recently published in MDM (e.g., [30], [65], [73]) and other major database and spatial conferences including SIGMOD, VLDB, ICDE and SIGSPATIAL. We expect the seminar to help the audience in identifying the possible future work in this intersection area. It can also be very beneficial for graduate students who search for PhD topics and research challenges. No prior knowledge is required to understand the spatial systems and approaches in the seminar. Yet, it requires basic machine learning knowledge, which is assumed to be there for the MDM audience. This seminar will act as an invitation to the mobile data management community to join arms for satisfying the emerging needs of big spatial data analysis and machine learning applications.

IV. Biographical Sketches

Ibrahim Sabek (PhD, University of Minnesota) is a Postdoctoral Associate at MIT. His research interests broadly include machine learning for systems, scalable data processing and querying, probabilistic databases, scalable knowledge base construction, and big spatial data management and analysis. Ibrahim has been named an NSF Computing Innovation Fellow (CIFellow) in 2020, and awarded the University of Minnesota Doctoral Dissertation Fellowship in 2019 for his dissertation focus on scalable machine learning for big spatial data and applications. His research work has won the first place of ACM SIGSPATIAL Student Research Competition (SRC) 2019, and has been nominated for the Best Paper Award of ACM SIGSPATIAL 2018. For more information, please visit: http://people.csail.mit.edu/ibrahimsabek/.

Mohamed F. Mokbel (PhD, Purdue University) is a Professor at University of Minnesota. His current research interests focus on building systems for big spatial data and applications. His research work has been recognized by the VLDB 10-years Best Paper Award, four conference Best Paper Awards, and the NSF CAREER Award. Mohamed is the past elected Chair of ACM SIGPATIAL, current Editor-in-Chief for Distributed and Parallel Databases Journal, and on the editorial board of ACM Books, ACM TODS, VLDB Journal, ACM TSAS, and GoeInformatica journals. He has also served as PC Vice Chair of ACM SIGMOD and PC Co-Chair for ACM SIGSPATIAL and IEEE MDM. Mohamed is an IEEE Fellow and an ACM Distinguished Scientist. For more information, please visit: www.cs.umn.edu/∼mokbel.

References

[1] A. Al-Mamun, H. Wu, and W. G. Aref. A Tutorial on Learned Multi-Dimensional Indexes. In *SIGSPATIAL*, 2020.

[2] L. Anselin et al. GeoDa: An Introduction to Spatial Data Analysis. *Journal of Geographical Analysis*, 38(1):5–22, 2006.

[3] Azure Geo AI. https://azure.microsoft.com/en-us/blog/microsoft-and-esri-launch-geospatial-ai-on-azure/.

[4] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt. RoadTracer: Automatic Extraction of Road Networks from Aerial Images. In *CVPR*, 2018.

[5] S. Bhadange, A. Arora, and A. Bhattacharya. GARUDA: A System for Large-scale Mining of Statistically Significant Connected Subgraphs. *PVLDB*, 9(13):1449–1452, 2016.

[6] C.-R. Chen and U. T. Kartini. K-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data. *Journal of Energies*, 10(2), 2017.

[7] F. Chen, Z. Chen, S. Biswas, S. Lei, N. Ramakrishnan, and C.-T. Lu. Graph Convolutional Networks with Kalman Filtering for Traffic Prediction. In *SIGSPATIAL*, 2020.

[8] T. Condie, P. Mineiro, N. Polyzoti, and M. Weimer. Machine Learning for Big Data (Tutorial). In *SIGMOD*, 2013.

[9] M. L. Damiani, A. Acquaviva, F. Hachem, and M. Rossini. Learning Behavioral Representations of Human Mobility. In *SIGSPATIAL*, 2020.

[10] E. Diday. Spatial Classification. *Journal of Discrete Applied Mathematics*, 156(8):1271–1294, 2008.

[11] X. Ding et al. UlTraMan: A Unified Platform for Big Trajectory Data Management and Analytics. In *VLDB*, 2018.

[12] O. Dovrat, I. Lang, and S. Avidan. Learning to Sample. In *CVPR*, 2019.

[13] A. Eldawy and M. F. Mokbel. The Era of Big Spatial Data (Tutorial). *PVLDB*, 10(12):1992–1995, 2017.

[14] ESRI ArcGIS. www.esri.com/en-us/arcgis/about-arcgis/overview.

[15] ESRI Tools for Hive. github.com/Esri/spatial-framework-for-hadoop.

[16] M. Ester, H. Kriegel, et al. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *SIGKDD*, 1996.

[17] J. H. Faghmous and V. Kumar. *Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities*, pages 83–116. Springer, 2014.

[18] R. Finkel and J. Bentley. Quad Trees a Data Structure for Retrieval on Composite Keys. *Acta Informatica*, 1974.

[19] R. Frank, M. Ester, and A. Knobbe. A Multi-relational Approach to Spatial Classification. In *SIGKDD*, 2009.

[20] R. Guhaniyogi and S. Banerjee. Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets. *Journal of Technometrics*, 60(4):430–444, 2018.

[21] C. Guo, B. Yang, J. Hu, and C. Jensen. Learning to Route with Sparse Trajectory Sets. In *ICDE*, 2018.

[22] T. Guo, K. Feng, et al. Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration. In *SIGMOD*, 2018.
[23] C. Hajdu and ron Ballagi. Towards a Quadtree based Approach to Learn Local Plans in Robotic Motion Planning. In *IEEE GPMC*, 2020.
[24] J. M. Hellerstein, C. R'e, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, and A. Kumar. The MADlib Analytics Library: or MAD Skills, the SQL. *PVLDB*, 5(12):1700–1711, 2012.
[25] J. Hu, C. Guo, B. Yang, et al. Stochastic Weight Completion for Road Networks Using Graph Convolutional Networks. In *ICDE*, 2019.
[26] J. Hughes et al. GeoMesa: A Distributed Architecture for Spatio-temporal Fusion. In *SPIE Defense+Security*, 2015.
[27] P. K. Jayaraman et al. Quadtree Convolutional Neural Networks. In *ECCV*, 2018.
[28] Z. Jiang, Y. Li, S. Shekhar, L. Rampi, and J. Knight. Spatial Ensemble Learning for Heterogeneous Geographic Data with Class Ambiguity: A Summary of Results. In *SIGSPATIAL*, 2017.
[29] Z. Jiang and S. Shekhar. *Spatial Big Data Science: Classification Techniques for Earth Observation Imagery*. Springer Publishing Company, 1st edition, 2017.
[30] M. Y. Kabir, S. Gruzdev, and S. Madria. STIMULATE: A System for Real-time Information Acquisition and Learning for Disaster Management. In *MDM*, 2020.
[31] L. J. Klein et al. PAIRS: A Scalable Geo-spatial Data Analytics Platform. In *IEEE Big Data*, 2015.
[32] M. Koubarakis et al. TELEIOS: A Database-powered Virtual Earth Observatory. In *VLDB*, 2012.
[33] T. Kraska. Learned Data Structures and Algorithms (Tutorial). In *SIGMOD*, 2019.
[34] A. Kumar et al. Data Management in Machine Learning: Challenges, Techniques, and Systems (Tutorial). In *SIGMOD*, 2017.
[35] N. Levine. *CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents*, pages 381–388. Springer, 2017.
[36] A. Li, S. Wang, W. Li, S. Liu, and S. Zhang. Predicting Human Mobility with Federated Learning. In *SIGSPATIAL*, 2020.
[37] P. Li, H. Lu, Q. Zheng, L. Yang, and G. Pan. LISA: A Learned Index Structure for Spatial Data. In *SIGMOD*, 2020.
[38] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *ICLR*, 2018.
[39] Y. Lin et al. Exploiting Spatiotemporal Patterns for Accurate Air Quality Forecasting Using Deep Learning. In *SIGSPATIAL*, 2018.
[40] Q. LIU, Y. Shen, and L. Chen. LHist: Towards Learning Multi-dimensional Histogram for Massive Spatial Data. In *ICDE*, 2021.
[41] Z. Lu, D. Agarwal, and I. S. Dhillon. A Spatio-temporal Approach to Collaborative Filtering. In *RecSys*, 2009.
[42] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou. LC-RNN: A Deep Learning Model for Traffic Speed Prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2018.
[43] Magellan: Geospatial analytics using spark. https://github.com/harsha2010/magellan.
[44] X. Meng et al. MLlib: Machine Learning in Apache Spark. *Journal of Machine Learning Research*, 17(1), 2016.
[45] R. T. Ng and J. Han. Efficient and Effective Clustering Methods for Spatial Data Mining. In *VLDB*, pages 144–155, 1994.
[46] V. Pandey, A. van Renen, A. Kipf, I. Sabek, J. Ding, and A. Kemper. The Case for Learned Spatial Indexes. In *AIDB@VLDB*, 2020.
[47] Y. Pang, K. Tsubouchi, T. Yabe, and Y. Sekimoto. Intercity Simulation of Human Mobility at Rare Events via Reinforcement Learning. In *SIGSPATIAL*, 2020.
[48] M. M. Patil and A. Magdy. LATEST: Learning-Assisted Selectivity Estimation Over Spatio-Textual Streams. In *ICDE*, 2021.
[49] T. Plötz et al. Neural Nearest Neighbors Networks. In *NIPS*, 2018.
[50] PostGIS. http://postgis.net/.
[51] PostgreSQL. https://www.postgresql.org/, 2019.
[52] J. Qi, G. Liu, C. S. Jensen, and L. Kulik. Effectively Learning Spatial Indices. In *VLDB*, 2020.
[53] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré. HoloClean: Holistic Data Repairs with Probabilistic Inference. *PVLDB*, 10(11):1190–1201, 2017.
[54] S. Rey et al. *PySAL: A Python Library of Spatial Analytical Methods*, pages 175–193. Springer, 2010.
[55] I. Sabek. Adopting Markov Logic Networks for Big Spatial Data and Applications. In *VLDB PhD Workshop*, 2019.
[56] I. Sabek and M. F. Mokbel. Machine Learning Meets Big Spatial Data (Tutorial). In *VLDB*, 2019.
[57] I. Sabek and M. F. Mokbel. Machine Learning Meets Big Spatial Data (Tutorial). In *ICDE*, 2020.
[58] I. Sabek and M. F. Mokbel. Sya: Enabling Spatial Awareness inside Probabilistic Knowledge Base Construction. In *ICDE*, 2020.
[59] I. Sabek, M. Musleh, and M. Mokbel. TurboReg: A Framework for Scaling Up Spatial Logistic Regression Models. In *SIGSPATIAL*, 2018.
[60] I. Sabek, M. Musleh, and M. F. Mokbel. A Demonstration of Sya: A Spatial Probabilistic Knowledge Base Construction System. In *SIGMOD*, pages 1689–1692, 2018.
[61] I. Sabek, M. Musleh, and M. F. Mokbel. Flash in Action: Scalable Spatial Data Analysis Using Markov Logic Networks. *PVLDB*, 12(12):1834–1837, 2019.
[62] Y.-L. K. Samo and S. Roberts. Scalable Nonparametric Bayesian Inference on Point Processes with Gaussian Processes. In *ICML*, 2015.
[63] A. D. Sarma, H. Lee, H. Gonzalez, et al. Efficient Spatial Sampling of Large Geographical Tables. In *SIGMOD*, 2012.
[64] M. Sethi, Y. Yan, A. Rangarajan, R. R. Vatsavai, and S. Ranka. Scalable Machine Learning Approaches for Neighborhood Classification Using Very High Resolution Remote Sensing Imagery. In *SIGKDD*, 2015.
[65] S. Shams, S. Goswami, and K. Lee. Deep Learning-Based Spatial Analytics for Disaster-Related Tweets: An Experimental Study. In *MDM*, 2019.
[66] S. Shekhar, C.-T. Lu, and P. Zhang. Detecting Graph-based Spatial Outliers: Algorithms and Applications (a Summary of Results). In *SIGKDD*, 2001.
[67] J. Shin, S. Wu, F. Wang, et al. Incremental Knowledge Base Construction Using DeepDive. *PVLDB*, 8(11):1310–1321, 2015.
[68] A. Shokry, M. Torki, and M. Youssef. DeepLoc: A Ubiquitous Accurate and Low-overhead Outdoor Cellular Localization System. In *SIGSPATIAL*, 2018.
[69] R. Stanojevic, S. Abbar, and M. Mokbel. W-edge: Weighing the Edges of the Road Network. In *SIGSPATIAL*, 2018.
[70] C. R. Stephens, V. Snchez-Cordero, and C. G. Salazar. Bayesian Inference of Ecological Interactions from Spatial Data. *Journal of Entropy*, 19(12), 2017.
[71] R. Tibshirani and P. Wang. Spatial Smoothing and Hot Spot Detection for CGH Data Using the Fused Lasso. *Biostatistics*, 9(1):18–29, 2008.
[72] T. VoPham et al. Emerging Trends in Geospatial Artificial Intelligence (geoAI): Potential Applications for Environmental Epidemiology. *Environmental Health*, 2018.
[73] H. Wang and H. Su. STAR: A Concise Deep Learning Framework for Citywide Human Mobility Prediction. In *MDM*, 2019.
[74] L. Wang, R. Christensen, F. Li, and K. Yi. Spatial Online Sampling and Aggregation. *PVLDB*, 9(3):84–95, 2015.
[75] W. Wang, J. Yang, and R. R. Muntz. STING: A Statistical Information Grid Approach to Spatial Data Mining. In *VLDB*, 1997.
[76] H. Wei, G. Zheng, H. Yao, and Z. Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *SIGKDD*, 2018.
[77] D. Xie, F. Li, B. Yao, G. Li, L. Zhou, and M. Guo. Simba: Efficient In-Memory Spatial Analytics. In *SIGMOD*, 2016.
[78] Y. Xie et al. An Unsupervised Augmentation Framework for Deep Learning Based Geospatial Object Detection: A Summary of Results. In *SIGSPATIAL*, 2018.
[79] H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou. Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation. *TKDE*, 29(11), 2017.
[80] J. Yu, Z. Zhang, and M. Sarwat. Spatial Data Management in Apache Spark: The GeoSpark Perspective and Beyond. *Journal of GeoInformatica*, pages 1–44, 2018.
[81] E. T. Zacharatou, A. Kipf, I. Sabek, et al. The Case for Distance-Bounded Spatial Approximations. In *CIDR*, 2021.
[82] J. Zhang, Y. Zheng, and D. Qi. Deep Spatio-temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI*, 2017.
[83] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *SIGMOD*, 1996.
[84] G. Zheng, S. L. Brantley, T. Lauvaux, and Z. Li. Contextual Spatial Outlier Detection with Metric Learning. In *SIGKDD*, 2017.