Results from Prior Support – Shashi Shekhar

Closest grant for Shashi Shekhar was the NSF/*CISE Expeditions in Computing: Understanding Climate Change: A Data Driven Approach* (Grant No. 1029711, $6.4M (approx.), 08/2010-08/2015). This grant supported two graduate students (50%) and one summer month faculty salary each year.

Intellectual Merits: Results were published in 8 peer-reviewed articles [1-9]. Space limitation does not permit a description of all the results. Thus, we summarize a few illustrative examples next.

- Spatial decision tree for geographical classification [3,4,6]: Given a raster spatial framework, as well as training and test sets, the spatial decision-tree learning (SDTL) problem aims to find a decision tree model that minimizes classification errors as well as salt-and-pepper noise. The SDTL problem is important due to many societal applications such as land cover classification in remote sensing. However, the SDTL problem is challenging due to the spatial autocorrelation of class labels, and the potentially exponential number of candidate trees. Related work is limited due to the use of local-test-based decision nodes, which cannot adequately model spatial autocorrelation during test phase, leading to high salt-and-pepper noise. In contrast, we proposed a focal-test-based spatial decision tree (FTSDT) model, where the tree traversal direction for a location is based on not only local but also focal (i.e., neighborhood) properties of the location. Experimental results on real world remote sensing datasets showed that the proposed approach reduced salt-and-pepper noise and improved classification accuracy. Preliminary results were published at the *IEEE International Conference on Data Mining* [3,4] and complete results appeared in the *IEEE Transactions on Data and Knowledge Engineering* [6].
- An inter-disciplinary survey on change footprint patterns and discovery techniques [5]: Given a change definition and a dataset about spatiotemporal (ST) phenomena, ST change footprint discovery is the process of identifying the location and/or time of such changes from the dataset. Change footprint discovery is fundamentally important for the study of climate change, the tracking of disease, and many other applications. Methods for detecting change footprints have emerged from a diverse set of research areas, ranging from time series analysis and remote sensing to spatial statistics. Researchers have much to learn from one another, but are stymied by inconsistent use of terminology and varied definitions of change across disciplines. Existing reviews focus on discovery methods for only one or a few types of change footprints (e.g., point change in a time series). To facilitate sharing of insights across disciplines, we conducted a multi-disciplinary review of ST change patterns and their respective discovery methods. We developed a taxonomy of possible ST change footprints and classified our review findings accordingly. This work allowed us to identify gaps in the research that we consider ripe for exploration, most notably change pattern discovery in vector ST datasets. In addition, we illustrate such pattern discovery using two case studies from historical GIS. A paper summarizing the survey has been published at the *Wiley Interdisciplinary Review: Data Mining and Knowledge Discovery* [5].
- Interesting sub-path discovery [1,2]: Given a spatiotemporal (ST) dataset and a path in its embedding spatiotemporal framework, the goal of the interesting sub-path discovery problem is to identify all interesting sub-paths defined by an interest measure. An important application domain of sub-path discovery is climate change, e.g., finding time periods of precipitation decreases. This problem is challenging due to the massive volume of data, the varying length of sub-paths and non-monotonicity of interestingness. Previous approaches find interesting unit sub-paths (e.g., unit time interval) or interesting points. In contrast, our work finds long intervals or sub-paths with arbitrary length. We model the computational structure of this problem as a Grid-based Directed Acyclic Graph (G-DAG). We propose a novel algorithm, namely, the Row-wise Traversal (after leaf-evaluation) with Column Pruning (RTCP) which brings dramatically down the memory cost for G-DAG traversal in the preliminary approach while also reducing CPU cost. Experimental evaluations on both synthetic and real datasets show that the RTCP algorithm is always the fastest in computational time among all the proposed algorithms. Results were published at ACM SIGSPATIAL Conference on GIS [1,2].

The results from our recent project not only advance spatial data science but also advance data science and computer science for a very broad range of data sets. The interdisciplinary survey [5, 7] we developed provided a way to bring together research ideas from various domains and enabled cross-fertilization of techniques. Researchers in need of finding a suitable change detection method may benefit from our work by learning techniques developed by other domains and further improve them for their own applications.

Broader Impacts: Two doctoral students, namely, X. Zhou and Z. Jiang, supported by this project defended their Ph.D. theses and joined the faculty of University of Iowa in Summer 2014 and University of Alabama in Summer 2016 respectively. We co-organized a special issue of IEEE Computing in Science and Engineering magazine [9], and co-authored two survey papers [4,7] and a vision paper [8]. In addition, Shekhar recorded and taught a massively open online course titled "From GPS and Google Maps to Spatial Computing" (Coursera, Fall 2014), which atteacted over 21,000 student across 182 countries. The video lectures from this course are now available on youtube.com for public access and are being used by students and researchers around the country. Shekhar also presented several invited lectures and keynotes at major conferences. Using the results obtained in our project, we enhanced the curriculum of two courses titled "Spatial Computing" (Csci 5715)" and "Spatial databases" (CSci 8715). Finally, the projects have engaged underrepresented students in research and education. In summer 2012, Ms. Rahni Sumler, an undergraduate student from North Carolina A&T University (a HBCU) had a six-week internship on spatiotemporal data analytics for understanding climate change.

Relationship with proposed work: The current research in the above project assumes geometry-specified shapes (e.g., circles, rectangles) of change footprints, which may not reflect the constraints in the natural environment (e.g., terrain, landscape). The new project will pursue physics-aware relationship and pattern mining techniques which automatically and efficiently extract/learn geography-specified constraints to guide the discovery process for change footprints and other spatio-temporal patterns.

REFERENCES

[1] Xun Zhou, Shashi Shekhar, Pradeep Mohan, Stefan Liess, and Peter K Snyder. Discovering Interesting Sub-paths in Spatiotemporal Datasets: A Summary of Results. In *Proc. the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 44–53. ACM, 2011.

[2] Xun Zhou, Shashi Shekhar, and Dev Oliver. Discovering persistent change windows in spatiotemporal datasets: A summary of results. In *Proc. 2nd ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (BigSpatial-2013)*, Orlando, FL, Nov. 5, 2013. ACM. (Best Paper Award).

[3] Zhe Jiang, Shashi Shekhar, Xun Zhou, Joseph Knight, and Jennifer Corcoran. Focal-test-based spatial decision tree learning: A summary of results. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. IEEE, 2013.

[4] Z. Jiang, S. Shekhar, A. Kamzin, and J. Knight, Learning a Spatial Ensemble of Classifiers for Raster Classification: A Summary of Results. In *2014 IEEE International Conference on Data Mining Workshop*, pages 15-18, 2014.

[5] X. Zhou, S. Shekhar, and R. Ali. Spatiotemporal Change Footprint Pattern Discovery: An Inter-disciplinary Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1):1–23, 2014.

[6] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, and J. Corcoran. Focal-test-based spatial decision tree learning. *Knowledge and Data Engineering, IEEE Transactions on Knowledge and Data Engineering, 27(6):1547-1559*, 2015.

[7] Shashi Shekhar, Zhe Jiang, Reem Y. Ali, Emre Eftelioglu, Xun Tang, Venkata Gunturi, and Xun Zhou. Spatiotemporal data mining: A computational perspective. *ISPRS International Journal of Geo-Information,* 4(4): 2306-2338, 2015.

[8] J. Faghmous, V. Kumar, and S. Shekhar (Guest Editors), IEEE Computing in Science and Engineering (*Special issue on Computing in Climate*), 17(6), 2015.

[9] S Shekhar, S. Feiner, and W. Aref, *Spatial Computing*, Communications of the ACM (cover article with an accompanying 4-minute video), 59(1):72-81, 2016.