Grand challenges for the spatial information community

Leye Wang[1,2] and Ouri Wolfson[3]

[1]Key Lab of High Confidence Software Technologies (Peking University), Ministry of Education, China
[2]Department of Computer Science and Technology, Peking University, China
[3]University of Illinois at Chicago, Chicago, IL, USA

Abstract: The spatial information (SI) community has an opportunity to address major societal and scientific problems including public health, climate change, air pollution, transportation, and others. Beyond the significant contributions made by the SI community, more can be done by focusing the efforts of the community, and generalizing them. Focus can be achieved by an IMAGENET-like spatial information database and competition. Generalization can be achieved by solving spatio-temporal information problems in disciplines such as neuroscience, chemistry, biology, astronomy, and engineering.

1 Introduction

The spatial information (SI) research community has an opportunity to address some major societal and scientific problems. Public health and epidemiology is an example whose importance is demonstrated by the recent COVID-19 shock. Climate change and sustainability is another example which is critical in emergencies such as floods, fires, hurricanes, and tornadoes that are becoming more frequent and more intense, and are made more dangerous due to the interdependence among food, water, transportation, and energy. Climate change is a critical factor in gradual non-emergency developments such as sea-level rise and human migration. An additional example is air pollution, which has been identified by the World Health Organization (WHO) as the world's largest single environmental

health risk; in the year 2012 alone around 7 million people died—one in eight of total global deaths—as a result of air pollution exposure[1]. Transportation is also an area where spatial and spatio-temporal information systems can reduce congestion and increase safety and convenience.

Most of these problems are aggravated by increased urbanization. Another complicating factor is increased inequality, which puts the burden of these developments disproportionately on the disadvantaged socio-economic classes.

Facing these challenges, the SI community has made significant contributions to address them, as a scan of SI journals and conferences will demonstrate. However, much more can be done by focusing the efforts of the community, and promoting by inter-disciplinarity. Consider focus first. While data-driven studies using SI are booming, a key problem of the SI community is that papers use different datasets, have different objectives, and efforts are thus incomparable. And when a paper improves on existing methods, it is hard to demonstrate the improvement without re-implementing these methods. Another problem faced by the SI community is an unfulfilled potential of generalizing its work. Specifically, existing efforts are focused on geospatial applications, whereas spatial and spatio-temporal information is abundant and useful in other domains such as neuroscience, chemistry, biology, astronomy, and engineering (e.g., computer-aided design and manufacturing).

The rest of the paper is organized as follows. In Section 2 we briefly survey the contributions of the SI community. In Sections 3 and 4 we discuss the two challenges mentioned above, namely focus and generalization, respectively. In Section 5 we conclude the paper.

2 Existing contributions of the SI community

Before presenting grand challenges, we first briefly summarize exiting contributions of the SI community up to date.

- **Data Collection with Ubiquitous Techniques.** In the recent decade, the SI community has adopted ubiquitous sensing techniques, which significantly improve the capability and scope of the geospatial applications [26]. The ubiquitous data collection sources include wireless sensors [20], social media [9], mobile crowdsensing [25], spatial crowdsourcing [13], etc.
- **Geospatial Data Mining with Artificial Intelligence.** Geospatial artificial intelligence, namely, GeoAI, has attracted tremendous interest from the SI community, with the booming of AI techniques in recent years. AI techniques have been applied widely in geospatial image processing, textual place name extraction, trajectory mining, etc. This led to the amazing prediction and modeling power of new AI-enhanced geospatial applications [12].
- **Streaming Spatial Data Processing.** As ubiquitous spatial data are collected continuously, some SI researchers focus on streaming spatial data, such as spatio-temporal [21] and geo-social queries [1]. It is worth noting that some off-the-shelf products have been upgraded with new features to facilitate streaming spatial data processing, such as Microsoft SQL Server [14].
- **Heterogeneous Spatial Data Integration.** As the modalities of spatial data are diverse, how to integrate or fuse multi-source heterogeneous data into a unified model

---

[1]https://www.who.int/mediacentre/news/releases/2014/air-pollution/en/

for the target application is always a key research direction in the SI community [5]. Researchers have also published multi-source integrated datasets for some target applications [22], or specific geographic areas [3] in order to inspire further research.

- **Spatial Big Data Infrastructure.** In addition to the data mining and AI algorithms, the SI community has also developed big data infrastructures [19], such as Spatial-Hadoop [8] and JUST [18], to support efficient storage and processing of various types of spatial big data generated by the ubiquitous sensing techniques.
- **Easy Data Access.** Another key achievement of the SI community in recent years is the emergence of more and more easily accessible datasets, with the promotion of open data activities in academia, industry, and governments. Common data sources include digital maps (e.g., OpenStreetMap [10]), call detail records (e.g., Milan [3]), transportation trips (e.g., NYC TLC Trip[2]), satellite and geospatial images (e.g., Google Street View [2]), etc. A notable example of the geospatial data warehouse is the governmental open data website like NYC Open Data Portal[3].

3 IMAGENET-like spatial information database and competition

The SI community may draw inspiration from others, e.g., computer vision (CV) and natural language processing, which faced the problem of incomparable results. The CV community has addressed this problem by the IMAGENET database and an annual competition in which approaches and results are comparable [6]. Particularly, IMAGENET has included 14 million images; the objects in each image are labeled by crowdsourced workers. The most famous competition based on the IMAGENET database is ILSVRC (IMAGENET Large Scale Visual Recognition Challenges)[4] which evaluates algorithms for object detection and image classification. Based on the IMAGENET database and the competitions, prestigious CV achievements in recent years have emerged, such as AlexNet [16] and ResNet [11].

Inspired by the success of IMAGENET, disparate efforts in the SI community may be rectified by an IMAGENET-like database that ensures replicability and enables comparison of results. However, directly copying the successful path of IMAGENET is infeasible for the SI community for the following reasons:

1. The data types of the SI community are much more diverse than in CV, where images are dominant. In SI research studies, data types include digital maps, satellite images, trajectories of moving objects, geo-tagged tweets, etc. With such diversity, it is challenging to design, build, and maintain a general-purpose database that can benefit the whole SI community. In fact, preliminary efforts of SI community have been devoted to create specific types of benchmark datasets, such as Movebank animal tracking data [15] and GeoLife human trajectory data [27], but the impacts are limited only to the related SI sub-area. A data warehouse consisting of diverse SI datasets is desired, so as to inspire cross-disciplinary and integrative ideas.

---

[2]https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
[3]https://opendata.cityofnewyork.us/
[4]http://image-net.org/challenges/LSVRC/

2. The scope of work of the SI community is broader and less focused than object identi-
fication in CV. It is thus impossible to define one or few focused SI challenges such as
ILSVRC that can drive the whole SI community. To this end, rather than a centralized-
style challenge proposed by the IMAGENET project, the SI community probably
needs to design a new distributed-style mechanism to encourage researchers in differ-
ent SI sub-areas to design an ILSVRC-like competition for their own sub-community.
This distributed organization would be more complicated to manage than the cen-
tralized one in IMAGENET, and thus how to promote it will need to be a whole
community effort. The effort may be seeded by the annual ACMGIS competition.
Particularly, the annual competition can be chosen to focus on a key topic of the SI
community for each year. In 2020, the competition may address public health and
epidemiology, with a focus on the spatial spread of infection and contact tracing. An-
other competition may address the impact of revolutionary developments in trans-
portation (mobility on demand, electrification, autonomous vehicles) on congestion
and air-pollution. Prediction of the impact of climate change under various scenarios
may be the subject of another competition.

4   Spatial information beyond geographic applications

Another key problem of the SI community is how to generalize its work beyond geospa-
tial applications. Actually, spatial and spatio-temporal information is common in many
research areas such as neuroscience, chemistry, biology, astronomy, and engineering (e.g.,
computer-aided design and manufacturing). We use the following examples to illustrate
the potential benefit of these disciplines from SI.

Neuroscience: The human brain is probably the most complex object in the universe.
It is also one of the least understood. For example, consider a basic human activity such
as sleep; there is no consensus concerning its purpose and how the purpose is achieved.
Higher level functions such as imagination, creativity, aggressiveness, and more generally
consciousness are highly controversial. Spatio-temporal data analysis can shed a light on
the relationship between the brain and abstract concepts such as mind and consciousness.

Furthermore, better understanding of the brain has Artificial Intelligence (AI) implica-
tions that go beyond medicine, biology, and philosophy. For example, the Artificial Neural
Network is a transformative computational paradigm inspired by the brain, but its ap-
plicability is currently limited to pattern recognition. Better understanding of the brain's
generation of common sense, understanding, and emotion can inform novel computational
and communication paradigms (see [17]).

Although the SI community has not done much research on the topic, the brain is fun-
damentally a spatial object. It has on the order of 100 billion neurons, each of which is
connected (via synapses) to about 1,000 other neurons on average [7]. Each brain-region
has a location in 3D space which is indexed using a common coordinate system, e.g., MNI
or Talairach [23]. Such a coordinate system is designed to be independent of individual
differences among humans. Furthermore, the brain also has a temporal component in the
sense that the neurons communicate by sending signals that travel from neuron to neuron
across time. Indeed, work on spatio-temporal approaches to brain understanding has been
published recently in the SI literature [24].

Chemistry: The chemical properties of molecules are often functions of their structure. An example is protein folding, in which the 3D structure determines function[5]. Furthermore, the folding occurs over time, thus this is a spatio-temporal process. The SI-chemistry joint study may generate new ideas toward grand societal challenges such as drug development.

Astronomy: Current SI research mostly focuses on geographic information, specifically on our planet. However, earth is only one planet in the universe whose diameter is estimated to be tens of billions light years; and furthermore, this universe continuously expands. In other words, the universe is a spatio-temporal object. Is it amenable to analysis using spatio-temporal tools developed by the SI community? If so, how, and if not, how should the tools be modified in order to facilitate such analysis [4]? Such research may help in answering fundamental questions in physics; for example, those related to dark matter, dark energy, and quantum effects.

5 Conclusion

In this vision paper we stated some preliminary ideas that we feel can significantly advance the SI community. We summarize the ideas as two keywords, "focus" and "generalization". By "focus" we mean identifying several key problems, e.g., traffic prediction, that are amenable to easy comparison of diverse and independent efforts. By "generalization", we mean that SI researchers should look beyond traditional geospatial applications and contribute to other disciplines such as neuroscience, chemistry, architecture and engineering, and astronomy.

Acknowledgments

References

[1] ALMASLUKH, A., AND MAGDY, A. Temporal geo-social personalized search over streaming data. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2019), pp. 189–198.

[2] ANGUELOV, D., DULONG, C., FILIP, D., FRUEH, C., LAFON, S., LYON, R., OGALE, A., VINCENT, L., AND WEAVER, J. Google street view: Capturing the world at street level. *Computer 43*, 6 (2010), 32–38.

[3] BARLACCHI, G., DE NADAI, M., LARCHER, R., CASELLA, A., CHITIC, C., TORRISI, G., ANTONELLI, F., VESPIGNANI, A., PENTLAND, A., AND LEPRI, B. A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data 2* (2015), 150055.

---

[5]https://en.wikipedia.org/wiki/Protein_folding

[4] BRAHEM, M., YEH, L., AND ZEITOUNI, K. Efficient astronomical query processing using spark. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2018), pp. 229–238.

[5] BUTENUTH, M., GÖSSELN, G. V., TIEDGE, M., HEIPKE, C., LIPECK, U., AND SESTER, M. Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry and Remote Sensing 62*, 5 (2007), 328–346.

[6] DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K., AND FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (2009), IEEE, pp. 248–255.

[7] DRACHMAN, D. A. *Do we have brain to spare?* AAN Enterprises, 2005.

[8] ELDAWY, A., AND MOKBEL, M. F. Spatialhadoop: A mapreduce framework for spatial data. In *2015 IEEE 31st international conference on Data Engineering* (2015), IEEE, pp. 1352–1363.

[9] HAHMANN, S., PURVES, R., AND BURGHARDT, D. Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes. *Journal of Spatial Information Science 2014*, 9 (2014), 1–36.

[10] HAKLAY, M., AND WEBER, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing 7*, 4 (2008), 12–18.

[11] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[12] HU, Y., GAO, S., LUNGA, D., LI, W., NEWSAM, S., AND BHADURI, B. GeoAI at ACM SIGSPATIAL: progress, challenges, and future directions. *SIGSPATIAL Special 11*, 2 (2019), 5–15.

[13] KAZEMI, L., AND SHAHABI, C. Geocrowd: enabling query answering with spatial crowdsourcing. In *Proceedings of the 20th international conference on advances in geographic information systems* (2012), pp. 189–198.

[14] KAZEMITABAR, S. J., DEMIRYUREK, U., ALI, M., AKDOGAN, A., AND SHAHABI, C. Geospatial stream query processing using Microsoft SQL Server StreamInsight. *Proceedings of the VLDB Endowment 3*, 1-2 (2010), 1537–1540.

[15] KRANSTAUBER, B., CAMERON, A., WEINZERL, R., FOUNTAIN, T., TILAK, S., WIKELSKI, M., AND KAYS, R. The Movebank data model for animal tracking. *Environmental Modelling & Software 26*, 6 (2011), 834–835.

[16] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[17] LAKE, B. M., ULLMAN, T. D., TENENBAUM, J. B., AND GERSHMAN, S. J. Building machines that learn and think like people. *Behavioral and brain sciences 40* (2017).

[18] LI, R., HE, H., WANG, R., HUANG, Y., LIU, J., RUAN, S., HE, T., BAO, J., AND ZHENG, Y. Just: Jd urban spatio-temporal data engine. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)* (2020), IEEE, pp. 1558–1569.

[19] LI, S., DRAGICEVIC, S., CASTRO, F. A., SESTER, M., WINTER, S., COLTEKIN, A., PETTIT, C., JIANG, B., HAWORTH, J., STEIN, A., ET AL. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS journal of Photogrammetry and Remote Sensing 115* (2016), 119–133.

[20] LIANG, S. H., CROITORU, A., AND TAO, C. V. A distributed geospatial infrastructure for sensor web. *Computers & Geosciences 31*, 2 (2005), 221–231.

[21] MAGDY, A., MOKBEL, M. F., ELNIKETY, S., NATH, S., AND HE, Y. Mercury: A memory-constrained spatio-temporal real-time search on microblogs. In *2014 IEEE 30th International Conference on Data Engineering* (2014), IEEE, pp. 172–183.

[22] MOOSAVI, S., SAMAVATIAN, M. H., PARTHASARATHY, S., TEODORESCU, R., AND RAMNATH, R. Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2019), pp. 33–42.

[23] TALAIRACH, J. Co-planar stereotaxic atlas of the human brain-3-dimensional proportional system. Georg Thieme Verlag, 1988.

[24] WOLFSON, O. Understanding the human brain via its spatio-temporal properties (vision paper). In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2018), pp. 85–88.

[25] ZHANG, D., WANG, L., XIONG, H., AND GUO, B. 4W1H in mobile crowd sensing. *IEEE Communications Magazine 52*, 8 (2014), 42–48.

[26] ZHENG, Y., CAPRA, L., WOLFSON, O., AND YANG, H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST) 5*, 3 (2014), 1–55.

[27] ZHENG, Y., XIE, X., MA, W.-Y., ET AL. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull. 33*, 2 (2010), 32–39.