RESEARCH STATEMENT

SHAOWEI LIN

1. Introduction

My dream is to develop architectural solutions for real-world challenges in smart cities, internet-of-things and robotics, by understanding how learning occurs within large systems. These systems behave, grow and evolve in response to their internal state and external environment, in a manner similar to the human body or an ant colony, where the components speak one complex language and the flow of information enhances the whole. This analogy reflects how I think about *learning* in systems. To build infrastructures that are robust and resource-efficient, I like to derive my strategies from mathematics, geometry, computation, statistics and biology. Besides developing solutions, I also hope to seed a culture where Big Data is democratized, by making information and machine intelligence accessible to all.

2. Challenges from Urbanization

Big Data and the Internet-of-Things have been in the research limelight for many years now, but the key driver behind these trends is Urbanization - increasingly large numbers of people living in a small area with a multitude of needs competing for limited resources. To maintain a high quality of life, it is crucial to keep people, businesses and governments informed, as well as to assist them in making wiser decisions. We also want to free them from mundane tasks, so that they can focus on more creative endeavors. Intelligent networks of machines play a critical role, and the following are three of the associated challenges.

2.1. **Higher-order Intelligence.** Through the Internet-of-Things, we have access to large volumes and varieties of sensors and actuators, including swarms of quadcopter drones and autonomous vehicles communicating through the network. In fact, the Internet-of-Things is itself a massive Robot; higher-order intelligence is needed to derive situation awareness and coordinate its components. Previously, handcrafted algorithms were employed to deal with each sitation, but an unsupervised approach is required today due to the system complexity. This malleable distributed intelligence learns autonomously over time to improve itself.

2.2. **Heterogeneity.** In Big Data analytics, it is estimated that 80% of the time is spent formatting and cleaning the raw data for more advanced processing. The same is true in the Internet-of-Things: much effort is required to get different devices and protocols to work with each other. Each protocol was created to fulfill a specific objective, so heterogeneity is the reality that future internet and data architectures have to wrestle with.

Date: May 12, 2015.

2.3. **Resource-constrained Networks.** Many network protocols and machine learning algorithms were designed for wired multi-core systems with an abundance of energy. However, in the Internet-of-Things such as wireless sensor networks, the devices are often battery or solar powered, have limited computational facilities, and communicate with other machines through noisy channels. Designing resource-efficient learning algorithms can significantly prolong the lifespan and reduce the overall cost of the wireless network. An effective architecture can also allow valuable resources to be commoditized and shared among different applications running on the network.

3. Strategies from Nature

Massive complex systems are common in nature. Many of them are extremely effective at adapting and thriving over time, and inspire new strategies for building learning systems. Below, I highlight my favorite examples from biology, mathematics and physics.

3.1. **Neural networks.** The nervous system in the human body is a prime example of an energy-efficient distributed online learning network. Firstly, while it does not transmit data losslessly along its channels, it preserves important salient features in a robust manner and facilitates the emergence of higher-order intelligence as the information flows. Compressive sensing and deep learning are two signal processing and machine learning strategies that mimic this behavior. Secondly, incoming signals are pushed up the hierarchical neural network, not pulled or polled by the data sink, which results in significant savings in energy. This observation explains why publish-subscribe protocols are becoming more popular than pull notifications in real-time sensing platforms. Thirdly, because information is processed and filtered in a distributed manner throughout the neural network, it is difficult to recover the initial sensed signals from the final brain activities. Hence, pushing intelligence to the edge of a sensor network can provide some limited but natural form of security and privacy.

3.2. **Type Theory.** Over the past century, what began as an alternative to set theory as a foundation for mathematics has become the inspiration for many powerful programming paradigms used in data analytics such as functional programming. In mathematics, there has been renewed interest in type theory because of its connections to radical developments in algebraic geometry and its key role in automated proof-checking. The ideas are simple: every object has a type, and every function is restricted to objects of a certain type. Some of these ideas are also found in semantic web and linked data circles, where the types show up as ontologies and provide interoperability in heterogeneous systems. Other concepts like immutability and lazy evaluation are also emerging in network and data architectures. Type theory can provide a clean separation between the logic behind a learning system and the implementation of this logic on networked resources.

3.3. **Singular Learning.** An important problem in statistical learning is analyzing the behavior of a learning algorithm for a given model when the set of training data grows large. In 2001, Sumio Watanabe made a breakthrough in this area when he discovered that the algorithm is adversely affected by singularities in the model and he proposed methods for overcoming these singularities. In simpler words, much like a black hole in astrophysics, the usual laws of learning break down near the singularity. Moreover, the commonly-used maximum likelihood method in machine learning has large generalization errors for singular statistical models such as neural networks. Today, deep learning overcomes this issue by adding sparsity to the model or by applying sampling techniques. A clearer picture of the underlying singular effects may be required to bring machine learning to the next level, possibly by introducing ideas from thermodynamics and statistical mechanics.

4. Current Work

4.1. **Algebraic Statistics.** Much of my early work was designing fast computational tools for analyzing singularities in statistical models, using ideal theoretic approaches in algebraic geometry. These tools can also be used for estimating integrals arising in information theory and Bayesian statistics. For instance, by using these algebraic methods, we resolved a long-standing question regarding the capacity of noncoherent SIMO channels [3], and quantified the volume of unfaithful regions that plague partial correlation testing [2]. These examples demonstrate how singular analysis is important in many practical areas of communications and machine learning.

4.2. **Deep Learning.** During my postdoctoral project with Berkeley and Stanford, I was intrigued by the unreasonable effectiveness of deep learning in many difficult problems. By introducing sparsity and training the model on natural images, the neurons learnt features which were discovered earlier in neuroscience experiments. Thus, neural networks perform well when the singular effects are accounted for. However, it is unclear if biological neurons accomplish this by enforcing sparsity or employing some dynamical schemes. I wondered also if it was possible to cure deep learning of its long training times and hyperparameter selection issues. Around that time, Chris Hillar from the Redwood Center for Theoretical Neuroscience introduced me to a new technique known as Minimum Probability Flow. The resulting training algorithm was simple, fast, distributed and resembled Hebbian learning in biological neural networks. Unfortunately, it was not clear how deep networks could be constructed to derive higher-order intelligence.

4.3. **Sensor Networks.** Back in Singapore, I joined the Sense and Sense-abilities program, an A*STAR national initiative to build a sensing and sense-making reference platform for smart cities. I believe that sensor networks form a part of the urban nervous system and that ideas from deep learning can aid us in their design. Six months later, I established the Sense-making Group whose primary role was developing machine learning techniques for improving the operation of wireless sensor networks (WSNs). The group strategy was to exploit low-complexity algorithms such as those from deep learning because power, storage, computation and communication are limited on WSNs. Because sensor networks measure many different modalities, each with different statistical properties, we designed an unsupervised algorithm that is able to learn suitable features for efficient data representation. Moreover, as data packets are frequently dropped due to noisy channels or low batteries, we wanted to train the algorithm on data sets with missing readings using a statistical and holistic approach. These constraints inspired us to use a sparse mean-field approximation of the conditional Restricted Boltzmann Machine that performed well on the Netflix data set. Our algorithm reduced prediction errors by 20% over state-of-the-art methods, despite being trained on temperature data where a quarter of the readings was missing [4]. We also successfully applied the same algorithm to data compression [1] and multimodal sensor fusion [5], showing that the same low-complexity algorithm is able to play many different important roles in WSNs.

4.4. **Data Analytics.** A secondary goal of the Sense-making Group was to derive actionable insights from real-time sensor data. One of our first projects was a collaboration with the National Environment Agency (NEA) on building large-scale WSNs for acoustic noise mapping in residential neighborhoods. Besides applying our missing data imputation algorithm and cubic spline interpolation for creating noise maps, we also used neural networks, decision trees and time series decomposition to detect outliers and identify noise sources such as construction sites, rush hour traffic and large bird populations. In another project with a high-tech farming company, we collected and analyzed greenhouse sensor readings and plant growth data to determine the best conditions required for cultivating different kinds of vegetables. Our results convinced the management to redesign their greenhouse for better ventilation and cooling, which led to a 20% increase in crop yields. As part of Singapore's Smart Nation initiatives, S&S worked closely with the Infocomm Development Authority (IDA) to provide technical input for the Jurong Lake District (JLD) multi-agency shared sensor network testbed. We conducted the network performance analysis and gas sensor profiling studies required to ensure system stability and data reliability. Our latest project is a collaboration with the Institute for High Performance Computing (IHPC) and the Housing Development Board (HDB) that involves developing a computational tool to simulate the interrelationships between the various urban microclimatic parameters on a modelling platform. Besides validating these simulations against actual sensor readings, we will also combine statistical and computational models in constructing more accurate real-time maps of the environmental conditions.

5. Future Plans

5.1. **Deep Probability Flow.** While deep learning has been largely successful in many difficult problems such as face recognition, speech recognition and natural language processing, the training of the neural network has to be performed offline as it is computationally expensive, often requiring several weeks or months. Application of the algorithm may also be challenging for inexperienced data scientists because of the many hyperparameters that usually need to be hand-picked. To mitigate these issues, I hope to develop a new approach for training deep neural networks by using ideas from Minimum Probability Flow. Some of the main benefits of this approach are as follows.

(1) The edge weights can be updated in an online fashion as data streams through the neural network. Through some form of Hebbian learning, the learning for each edge occurs locally, depending only on the values of the neurons they connect.

(2) Because of the distributed nature of this algorithm, parallel implementations can be written to work more efficiently on a cluster of GPUs. New computational chips, either organic or inorganic, could be designed to perform this kind of learning.
(3) Just as Google revolutionized internet search by indexing web pages efficiently, we can exploit neural networks in hashing other kinds of information such as speech, images, video and sensor data for more effective querying.
(4) This algorithm could be applied towards deep reinforcement learning, an area popularized by Demis Hassabis of Google DeepMind, so that complex systems can learn to optimize its operations gradually over time without much human supervision.

5.2. **Functional Web.** Ideas like 'named-data networking (NDN)' and 'moving the compute to the data' are gaining traction because traditional management models for network resources such as storage, computation and communications are gradually becoming irrelevant. For example, in bioinformatics, genome data is usually on the scale of petabytes so it makes more sense to run algorithms at the data source rather than to copy or move the data elsewhere. To accommodate these needs, I hope to design a data network architecture based on time-tested principles from type theory, functional programming, linked data and publish-subscribe protocols. As with NDN, this architecture could sit as a thin overlay on top of existing internet protocols or even serve as the base layer for other protocols. The primary goal is to create a strongly-typed network of immutable linked data objects and functions, where the data sets and function libraries may be cached at different physical locations in the network as the need arises. By allowing types and ontologies (collections of types) to grow organically over time, the architecture could alleviate the need for large-scale standardization for the Internet-of-Things.

References

[1] Mohammad Abu Alsheikh, Puay Kai Poh, Shaowei Lin, Hwee-Pink Tan, and Dusit Niyato. Efficient data compression with error bound guarantee in wireless sensor networks. In *Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pages 307–311. ACM, 2014.

[2] Shaowei Lin, Caroline Uhler, Bernd Sturmfels, and Peter Bühlmann. Hypersurfaces and their singularities in partial correlation testing. *Foundations of Computational Mathematics*, 14(5):1079–1116, 2014.

[3] Veniamin I Morgenshtern, Erwin Riegler, Wei Yang, Giuseppe Durisi, Shaowei Lin, Bernd Sturmfels, and H Bolcskei. Capacity pre-log of noncoherent simo channels via hironaka's theorem. *Information Theory, IEEE Transactions on*, 59(7):4213–4229, 2013.

[4] Liang Ze Wong, Huiling Chen, Shaowei Lin, and Daniel Chongli Chen. Imputing missing values in sensor networks using sparse data representations. In *Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pages 227–230. ACM, 2014.

[5] Wenyu Zhang, Xiaoping Ma, and Shaowei Lin. Multimodal fusion for sensor data using stacked autoencoders. *submitted*, 2014.

Institute for Infocomm Research
Current address: 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632
E-mail address: `lins@i2r.a-star.edu.sg`