Best Paper Award Abstracts

NIPS 2018

Safe and Nested Subgame Solving for Imperfect-Information Games

In imperfect-information games, the optimal strategy in a subgame may depend on the strategy in other, unreached subgames. Thus a subgame cannot be solved in isolation and must instead consider the strategy for the entire game as a whole, unlike perfect-information games. Nevertheless, it is possible to first approximate a solution for the whole game and then improve it in individual subgames. This is referred to as subgame solving. We introduce subgame-solving techniques that outperform prior methods both in theory and practice. We also show how to adapt them, and past subgame-solving techniques, to respond to opponent actions that are outside the original action abstraction; this significantly outperforms the prior state-of-the-art approach, action translation. Finally, we show that subgame solving can be repeated as the game progresses down the game tree, leading to far lower exploitability. These techniques were a key component of Libratus, the first AI to defeat top humans in heads-up no-limit Texas hold'em poker.

Variance-based Regularization with Convex Objectives

We develop an approach to risk minimization and stochastic optimization that provides a convex surrogate for variance, allowing near-optimal and computationally efficient trading between approximation and estimation error. Our approach builds off of techniques for distributionally robust optimization and Owen's empirical likelihood, and we provide a number of finite-sample and asymptotic results characterizing the theoretical performance of the estimator. In particular, we show that our procedure comes with certificates of optimality, achieving (in some scenarios) faster rates of convergence than empirical risk minimization by virtue of automatically balancing bias and variance. We give corroborating empirical evidence showing that in practice, the estimator indeed trades between variance and absolute performance on a training sample, improving out-of-sample (test) performance over standard empirical risk minimization for a number of classification problems.

EMNLP 2018

How Much Reading Does Reading Comprehension Require? A Critical Investigation of Popular Benchmarks

Many recent papers address reading comprehension, where examples consist of (question, passage, answer) tuples. Presumably, a model must combine information from both questions and passages to predict corresponding answers. However, despite intense interest in the topic, with hundreds of published papers vying for leaderboard dominance, basic questions about the difficulty of many popular benchmarks remain unanswered. In this paper, we establish sensible baselines for the bAbI, SQuAD, CBT, CNN, and Whodid-What datasets, finding that question- and passage-only models often perform surprisingly well. On 14 out of 20 bAbI tasks, passage-only models achieve greater than 50% accuracy, sometimes matching the full model. Interestingly, while CBT provides 20-sentence passages, only the last is needed for comparably accurate prediction. By comparison, SQuAD and CNN appear better-constructed.

Linguistically-Informed Self-Attention for Semantic Role Labeling

Current state-of-the-art semantic role labeling (SRL) uses a deep neural network with no explicit linguistic features. However, prior work has shown that gold syntax trees can dramatically improve SRL decoding, suggesting the possibility of increased accuracy from explicit modeling of syntax. In this work, we present linguistically-informed self-attention (LISA): a neural

network model that combines multi-head self-attention with multi-task learning across dependency parsing, part-ofspeech tagging, predicate detection and SRL. Unlike previous models which require significant pre-processing to prepare linguistic features, LISA can incorporate syntax using merely raw tokens as input, encoding the sequence only once to simultaneously perform parsing, predicate detection and role labeling for all predicates. Syntax is incorporated by training one attention head to attend to syntactic parents for each token. Moreover, if a high-quality syntactic parse is already available, it can be beneficially injected at test time without re-training our SRL model. In experiments on CoNLL-2005 SRL, LISA achieves new state-of-the-art performance for a model using predicted predicates and standard word embeddings, attaining 2.5 F1 absolute higher than the previous state-of-the-art on newswire and more than 3.5 F1 on outof-domain data, nearly 10% reduction in error. On ConLL-2012 English SRL we also show an improvement of more than 2.5 F1. LISA also out-performs the state-of-the-art with contextually-encoded (ELMo) word representations, by nearly 1.0 F1 on news and more than 2.0 F1 on out-of-domain text.

OSDI 18

REPT: Reverse Debugging of Failures in Deployed Software

Debugging software failures in deployed systems is important because they impact real users and customers. However, debugging such failures is notoriously hard in practice because developers have to rely on limited information such as memory dumps. The execution history is usually unavailable because high-fidelity program tracing is not affordable in deployed systems.

In this paper, we present REPT, a practical system that enables reverse debugging of software failures in deployed systems. REPT reconstructs the execution history with high fidelity by combining online lightweight hardware tracing of a program's control flow with offline binary analysis that recovers its data flow. It is seemingly impossible to recover data values thousands of instructions before the failure due to information loss and concurrent execution. REPT tackles these challenges by constructing a partial execution order based on timestamps logged by hardware and iteratively performing forward and backward execution with error correction.

We design and implement REPT, deploy it on Microsoft Windows, and integrate it into Windows Debugger. We evaluate REPT on 16 real-world bugs and show that it can recover data values accurately (92% on average) and efficiently (less than 20 seconds) for these bugs. We also show that it enables effective reverse debugging for 14 bugs.

LegoOS: A Disseminated, Distributed OS for Hardware Resource Disaggregation

The monolithic server model where a server is the unit of deployment, operation, and failure is meeting its limits in the face of several recent hardware and application trends. To improve heterogeneity, elasticity, resource utilization, and failure handling in datacenters, we believe that datacenters should break monolithic servers into disaggregated, network-attached hardware components. Despite the promising benefits of hardware resource disaggregation, no existing OSes or software systems can properly manage it. We propose a new OS model called the splitkernel to manage disaggregated systems. Splitkernel disseminates traditional OS functionalities into loosely-coupled monitors, each of which runs on and manages a hardware component. Using the splitkernel model, we built LegoOS, a new OS designed for hardware resource disaggregation. LegoOS appears to users as a set of distributed servers. Internally, LegoOS cleanly separates processor, memory, and storage devices both at the hardware level and the OS level. We implemented LegoOS from scratch and evaluated it by emulating hardware

components using commodity servers. Our evaluation results show that LegoOS's performance is comparable to monolithic Linux servers, while largely improving resource packing and failure rate over monolithic clusters.

Crypto 2018

Yes, There is an Oblivious RAM Lower Bound!

An Oblivious RAM (ORAM) introduced by Goldreich and Ostrovsky [JACM'96] is a (possibly randomized) RAM, for which the memory access pattern reveals no information about the operations performed. The main performance metric of an ORAM is the bandwidth overhead, i.e., the multiplicative factor extra memory blocks that must be accessed to hide the operation sequence. In their seminal paper introducing the ORAM, Goldreich and Ostrovsky proved an amortized $$\varOmega (\lg n)$$ bandwidth overhead lower bound for ORAMs with memory size n. Their lower bound is very strong in the sense that it applies to the "offline" setting in which the ORAM knows the entire sequence of operations ahead of time. However, as pointed out by Boyle and Naor [ITCS'16] in the paper "Is there an oblivious RAM lower bound?", there are two caveats with the lower bound of Goldreich and Ostrovsky: (1) it only applies to "balls in bins" algorithms, i.e., algorithms where the ORAM may only shuffle blocks around and not apply any sophisticated encoding of the data, and (2), it only applies to statistically secure constructions. Boyle and Naor showed that removing the "balls in bins" assumption would result in super linear lower bounds for sorting circuits, a long standing open problem in circuit complexity. As a way to circumventing this barrier, they also proposed a notion of an "online" ORAM, which is an ORAM that remains secure even if the operations arrive in an online manner. They argued that most known ORAM constructions work in the online setting as well. Our contribution is an $$\varOmega (\lg n)$$ lower bound on the bandwidth overhead of any online ORAM, even if we require only computational security and allow arbitrary representations of data, thus greatly strengthening the lower bound of Goldreich and Ostrovsky in the online setting. Our lower bound applies to ORAMs with memory size n and any word size $$r \ge 1$$. The bound therefore asymptotically matches the known upper bounds when $$r = \varOmega (\lg ^2 n)$$.

Multi-Theorem Preprocessing NIZKs from Lattices

Non-interactive zero-knowledge (NIZK) proofs are fundamental to modern cryptography. Numerous NIZK constructions are known in both the random oracle and the common reference string (CRS) models. In the CRS model, there exist constructions from several classes of cryptographic assumptions such as trapdoor permutations, pairings, and indistinguishability obfuscation. Notably absent from this list, however, are constructions from standard lattice assumptions. While there has been partial progress in realizing NIZKs from lattices for specific languages, constructing NIZK proofs (and arguments) for all of $$\mathsf {NP}$$ from standard lattice assumptions remains open. In this work, we make progress on this problem by giving the first construction of a multi-theorem NIZK argument for $$\mathsf {NP}$$ from standard lattice assumptions in the preprocessing model. In the preprocessing model, a (trusted) setup algorithm generates proving and verification keys. The proving key is needed to construct proofs and the verification key is needed to check proofs. In the multi-theorem setting, the proving and verification keys should be reusable for an unbounded number of theorems without compromising soundness or zero-knowledge. Existing constructions of NIZKs in the preprocessing model (or even the designated-verifier model) that rely on weaker assumptions

like one-way functions or oblivious transfer are only secure in a single-theorem setting. Thus, constructing multi-theorem NIZKs in the preprocessing model does not seem to be inherently easier than constructing them in the CRS model. We begin by constructing a multi-theorem preprocessing NIZK directly from context-hiding homomorphic signatures. Then, we show how to efficiently implement the preprocessing step using a new cryptographic primitive called blind homomorphic signatures. This primitive may be of independent interest. Finally, we show how to leverage our new lattice-based preprocessing NIZKs to obtain new malicious-secure MPC protocols purely from standard lattice assumptions

INFOCOM 2018

Understanding Ethereum via Graph Analysis

Being the largest blockchain with the capability of running smart contracts, Ethereum has attracted wide attention and its market capitalization has reached 20 billion USD. Ethereum not only supports its cryptocurrency named Ether but also provides a decentralized platform to execute smart contracts in the Ethereum virtual machine. Although Ether's price is approaching 200 USD and nearly 600K smart contracts have been deployed to Ethereum, little is known about the characteristics of its users, smart contracts, and the relationships among them. To fill in the gap, in this paper, we conduct the first systematic study on Ethereum by leveraging graph analysis to characterize three major activities on Ethereum, namely money transfer, smart contract creation, and smart contract invocation. We design a new approach to collect all transaction data, construct three graphs from the data to characterize major activities, and discover new observations and insights from these graphs. Moreover, we propose new approaches based on cross-graph analysis to address two security issues in Ethereum. The evaluation through real cases demonstrates the effectiveness of our new approaches

WiFED: WiFi Friendly Energy Delivery with Distributed Beamforming

Wireless RF energy transfer for indoor sensors is an emerging paradigm that ensures continuous operation without battery limitations. However, high power radiation within the ISM band interferes with the packet reception for existing WiFi devices. The paper proposes the first effort in merging the RF energy transfer functions within a standards compliant 802.11 protocol to realize practical and WiFi-friendly Energy Delivery (WiFED). The WiFED architecture is composed of a centralized controller that coordinates the actions of multiple distributed energy transmitters (ETs), and a number of deployed sensors that periodically request energy from the ETs. The paper first describes the specific 802.11 supported protocol features that can be exploited by sensors to request energy and for the ETs to participate in the energy delivery process. Second, it devises a controller-driven bipartite matching-based algorithmic solution that assigns the appropriate number of ETs to energy requesting sensors for an efficient energy transfer process. The proposed in-band and protocol supported coexistence in WiFED is validated via simulations and partly in a software defined radio testbed, showing 15% improvement in network lifetime and 31% reduction in the charging delay compared to the classical nearest distance-based charging schemes that do not anticipate future energy needs of the sensors and are not designed to co-exist with wifi systems.