# Elias B. Khalil

## Contact Information

Klaus Advanced Computing Building
266 Ferst Dr. NW
Atlanta, GA 30332 USA

*Phone:* (404) 429 3015
*E-mail:* elias.khalil@cc.gatech.edu
*Webpage:* www.ekhalil.com

## Research Areas

discrete optimization, machine learning, integer programming, deep learning

## Education

**Georgia Institute of Technology**, Atlanta, Georgia, USA

Ph.D. Candidate, Computational Science & Engineering — **2014 – 2019 (expected)**

- Advisor: Bistra Dilkina
- Thesis: *Towards Tighter Integration of Machine Learning & Discrete Optimization*
- Committee: Bistra Dilkina, George Nemhauser, Shabbir Ahmed, Le Song, Tuomas Sandholm
- Minor area: Operations Research (School of Industrial & Systems Engineering)

M.S., Computer Science — **2012 – 2014**

- Thesis: *Optimizing the Structure of Diffusion Networks: Theory and Algorithms*
- Committee: Bistra Dilkina, Le Song, Duen Horng (Polo) Chau

**American University of Beirut (AUB)**, Beirut, Lebanon

B.S., Computer Science — **2009 – 2012**

- Final Project: *Optimized Summation of Polynomial Multiplications using Funnel Heaps*
- Dean's Honor List, 2009 – 2011

## Fellowships

IBM Ph.D. Fellowship ($30,000) — **2017 – 2018**
Awarded to exceptional Ph.D. students in a worldwide competitive process

Marshall D. Williamson Fellowship ($2,600), Georgia Institute of Technology — **2014**
Awarded to the top 2nd year Master's student at the College of Computing

Donald V. Jackson Fellowship ($1,500), Georgia Institute of Technology — **2013**
Awarded to the top 1st year Master's student at the College of Computing

Association Philippe Jabre Fellowship ($5,000) — **2012 – 2013**
Awarded to outstanding students in Lebanon to support graduate education abroad

## Paper & Poster Awards

First Prize, Poster Competition, INFORMS Annual Meeting — **2017**
*Machine Learning for Integer Programming*; Out of over 100 participants in all areas of operations research

Outstanding Poster Award, NemFest Workshop in Celebration of Nemhauser and Nemirovski — **2017**
*Learning to Run Heuristics in Tree Search*; Out of over 20 participants in all areas of optimization

Best Paper Award, NIPS Workshop on Frontiers of Network Analysis — **2013**
*CuttingEdge: Influence minimization in networks*; Out of over 20 participants; As Master's student

## Professional Experience

**Georgia Institute of Technology**, Atlanta, Georgia USA
*Graduate Research Assistant* — **August 2014 – Present**

**IBM Research AI**, Yorktown Heights, New York USA
*Research Intern – Automated Machine Learning & Data Science* **August 2017 – Dec. 2017**

**IBM Research**, Yorktown Heights, New York USA
*Research Intern* **May 2016 – July 2016**

**Symantec Corporation**, Culver City, California USA
*Research Intern, Research Labs* **May 2013 – August 2013**

## Preprints

[1] **Elias B. Khalil**, Amrita Gupta, Bistra Dilkina. (2018). Combinatorial Attacks on Binarized Neural Networks. In submission to the International Conference on Learning Representations (ICLR). arXiv:1810.03538 [cs.LG].

## Conference Publications

[2] **Elias B. Khalil***, Hanjun Dai* (*co-first authors), Yuyu Zhang, Bistra Dilkina, Le Song. (2017). Learning Combinatorial Optimization Algorithms over Graphs. Neural Information Processing Systems (NIPS). **Spotlight presentation, top 5% of submissions**.

[3] Afshar, Ardavan, Joyce C. Ho, Bistra Dilkina, Ioakeim Perros, **Elias B. Khalil**, Li Xiong, and Vaidy Sunderam. (2017). CP-Ortho: An orthogonal tensor factorization framework for spatio-temporal data. Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2017.

[4] **Elias B. Khalil**, Bistra Dilkina, George Nemhauser, Shabbir Ahmed, Yufen Shao. (2017). Learning to Run Heuristics in Tree Search. Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI).

[5] Fatemeh Nargesian, Udayan Khurana, Horst Samulowitz, **Elias B. Khalil**, Deepak Turaga. (2017). Learning Feature Engineering for Classification. Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI).

[6] Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, **Elias B. Khalil**, Shuang Li, Le Song, Hongyuan Zha. (2017) Fake News Mitigation via Point Process Based Intervention. International Conference on Machine Learning (ICML).

[7] **Elias B. Khalil**, Pierre Le Bodic, Le Song, George Nemhauser, Bistra Dilkina. (2016). Learning to Branch in Mixed Integer Programming. 30th AAAI Conference on Artificial Intelligence (AAAI).

[8] **Elias B. Khalil**, Bistra Dilkina, Le Song. (2014). Scalable Diffusion-Aware Optimization of Network Topology. 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

## Journal Publications

[9] Wenwen Zhang, Subhrajit Guhathakurta, **Elias B. Khalil**. (2018). The impact of private autonomous vehicles on vehicle ownership and unoccupied VMT generation. Transportation Research Part C: Emerging Technologies.

[10] Acar Tamersoy, **Elias B. Khalil**, Bo Xie, Stephen Lenkey, Brian Routledge, Duen Horng Chau, Shamkant Navathe. (2014). Large-scale insider trading analysis: patterns and discoveries. Social Network Analysis and Mining (SNAM), 4(1), pp. 1 – 17.

## Refereed Workshop or Short Papers

[11] **Elias B. Khalil**, Bistra Dilkina. (2018). Training Binary Neural Networks with Combinatorial Optimization. Extended Abstract. 15th International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR).

[12] Udayan Khurana, Fatemeh Nargesian, Horst Samulowitz, **Elias B. Khalil**, Deepak Turaga. (2016). Automating Feature Engineering. Workshop on Artificial Intelligence for Data Science at NIPS.

[13] **Elias B. Khalil**. (2016). Machine Learning for Integer Programming. Proceedings of the Doctoral Consortium at the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI).

[14] Sucheta Soundarajan, Acar Tamersoy, **Elias B. Khalil**, Tina Eliassi-Rad, Duen Horng Chau, Brian Gallagher, Kevin Roundy. (2016). Generating Graph Snapshots from Streaming Edge Data (poster paper). 25th International World Wide Web Conference (WWW).

[15] **Elias B. Khalil**, Bistra Dilkina, Le Song. (2013). CuttingEdge: Influence minimization in networks. Workshop on Frontiers of Network Analysis: Methods, Models, and Applications at (NIPS). **Best Paper award**.

## Selected Talks

1. INFORMS Annual Meeting — Phoenix, USA, November **2018**
2. International Symposium on Mathematical Programming — Bordeaux, France, July **2018**
3. CPAIOR Masterclass (**Invited Tutorial Speaker**) — Delf, The Netherlands, June **2018**
4. NIPS (**Spotlight talk, Travel award**) — Long Beach, USA, December **2017**
5. INFORMS Annual Meeting — Houston, USA, November **2017**
6. IJCAI (**Travel award**) — Melbourne, Australia, August **2017**
7. INFORMS Annual Meeting — Nashville, USA, November **2016**
8. IBM Research — Yorktown Heights, USA, June **2016**
9. INFORMS Optimization Society Conference — Princeton, USA, March **2016**
10. AAAI Conference on Artifical Intelligence (**Travel award**) — Phoenix, USA, February **2016**
11. International Symposium on Mathematical Programming — Pittsburgh, USA, July **2015**
12. Knowledge Discovery & Data Mining (KDD) — New York City, USA, August **2014**

## Selected Posters

1. Theoretical Foundation of Deep Learning workshop — Atlanta, USA, October **2018**
2. INFORMS Annual Meeting (**Best Poster**) — Houston, USA, November **2017**
3. Doctoral Consortium on Computational Sustainability — Los Angeles, USA, July **2017**
4. NemFest Workshop in Celebration of Nemhauser and Nemirovski (**Best Poster**) — Atlanta, USA, May **2017**
5. Doctoral Consortium at IJCAI — New York City, USA, July **2016**
6. Mixed Integer Programming Workshop (**Travel award**) — Chicago, USA, June **2015**
7. Georgia Tech Research and Innovation Conference (**Best Poster**) — Atlanta, USA, Feb. **2015**
8. NIPS Workshop: Frontiers of Network Analysis (**Best Paper**) — Lake Tahoe, USA, Dec. **2013**

## Travel Awards

| | |
|---|---|
| CPAIOR ($250) | **2018** |
| NIPS ($800) | **2017** |
| IJCAI ($1,000) | **2016** |
| AAAI ($125) | **2016** |
| Mixed Integer Programming Workshop ($500) | **2015** |
| Georgia Tech Career, Research and Innovation Conference ($1,500; twice) | **2015, 2016** |

## Teaching Experience

**Tutorial Presenter** **2018**
CPAIOR '18 Master Class on Machine Learning for Discrete Optimization
Delft, The Netherlands

**Teaching Assistant** **2014, 2018**
*Computational Science & Engineering Algorithms* (CSE 6140)
Georgia Institute of Technology, Atlanta, Georgia USA
– Fall 2018: Prof. Umit Catalyurek, 160 students
– Fall 2014: Prof. Bistra Dilkina, 90 students
– Gave multiple full lectures on approximation algorithms, local search, submodular optimization
– Helped design course assignments and projects

**Guest Lecturer** **2018**
*Topics in Discrete Optimization and Learning* (CSCI 699), Spring 2018
University of Southern California, Los Angeles, California, USA
– Contributed to the design of the course curriculum
– Gave a lecture on recent advances in deep reinforcement learning for optimization

**Mentor** **2013, 2015**
Georgia Institute of Technology, Atlanta, Georgia USA
– Sachin Grover: Undergraduate in Computer Science at IIT, Jodhpur
Summer research internship, Summer 2015: "Online Learning in Branch-and-Bound"
Currently Ph.D. Student in Computer Science, Carnegie Mellon University
– Samuel Clarke: Undergraduate in Computer Science at Georgia Tech
Independent research under Prof. Polo Chau, 2013: "Graph Mining with SQLite"
Currently M.S. Student in Robotics, Carnegie Mellon University

## Academic Service

**Program Committee member**
ICLR: International Conference on Learning Representations **2019**
AISTATS: International Conference on Artificial Intelligence and Statistics **2019**
AAAI Conference on Artificial Intelligence **2017, 2018**
NIPS: Neural Information Processing Systems **(Top 30% of reviewers in 2018)** **2017, 2018**
ICML: International Conference on Machine Learning **2018**

**Journal Reviewer**
Operations Research
INFORMS Journal of Computing
INFORMS Journal of Optimization
Annals of Operations Research
Computers & Operations Research
Journal of Machine Learning Research (JMLR)
IEEE Transactions on Knowledge and Data Engineering (TKDE)

**Conference Reviewer**
AAAI (2015, 2016), Constraint Programming (2016), IJCAI (2016), KDD (2015, 2016)

**Vice President, Graduate Student Association**
Computational Science & Engineering, Georgia Tech **2016 – 2018**
– Organized HotCSE student seminar (25 talks)

- Organized student interviews with 15 faculty candidates
- Led CSE soccer teams in four Georgia Tech intramurals tournaments

## Grant Writing Experience

**Office of Naval Research**: *Integrating Machine Learning and Integer Programming* **2018**
Assisted Bistra Dilkina, George Nemhauser, Sebastian Pokutta.
Submitted in September 2018 (call).

**ExxonMobil**: *Leveraging Machine Learning and High-Performance Computing for Mixed Integer Programming* **2016**
Assisted Shabbir Ahmed, David Bader, Bistra Dilkina, George Nemhauser.
Granted, Dec. 2016 – Dec. 2017 ($405,000).

## Patents Granted

*Systems and Methods for Adjusting Suspiciousness Scores in Event-Correlation Graphs* **2015**
While at Symantec. Filed in 2013, Granted in 2015. US9148441 B1

*Systems and Methods for Using Event-Correlation Graphs to Detect Attacks on Computing Systems* **2015**
While at Symantec. Filed in 2013, Granted in 2015. US9141790 B2

## References

**Bistra Dilkina**
WiSE Gabilan Assistant Professor
Department of Computer Science
University of Southern California, Los Angeles, California, USA
Email: dilkina@usc.edu

**George Nemhauser**
A. Russell Chandler III Chair and Institute Professor
School of Industrial & Systems Engineering
Georgia Institute of Technology, Atlanta, Georgia USA
Email: gn3@gatech.edu

**Tuomas Sandholm**
Angel Jordan Professor of Computer Science
Computer Science Department
Carnegie Mellon University, Pittsburgh, PA USA
Email for recommendation letters: jpacker@andrew.cmu.edu, Jessica Packer (Assistant to T. Sandhom)
Email: sandholm@cs.cmu.edu

**Shabbir Ahmed**
Anderson-Interface Chair and Professor
School of Industrial & Systems Engineering
Georgia Institute of Technology, Atlanta, Georgia USA
Email: sahmed@isye.gatech.edu

**Andrea Lodi**
Canada Excellence Research Chair in Data Science for Real-Time Decision-Making and Professor
Department of Mathematical and Industrial Engineering
Polytechnique Montréal, Canada
Email: andrea.lodi@polymtl.ca

# Research Statement

Elias B. Khalil
www.ekhalil.com

Discrete optimization solvers have evolved considerably over the past two decades, and are now capable of handling problems with hundreds of thousands of variables and constraints. Simultaneously, new decision-making tasks have been identified in emerging domains. For instance, the advent of autonomous vehicles and the new field of "computational sustainability" have generated discrete optimization problems of high societal priority that must be addressed. These problems have peculiar combinatorial structure, one that solvers may not be well-equipped to handle. Another characteristic of such upsurging applications is that they produce optimization problems at a large velocity (e.g. hourly planning for a fleet of autonomous vehicles). Towards achieving the next giant leap in the performance of discrete optimization solvers, these challenges – new combinatorial structures and high problem velocity – must be met.

My research has focused on addressing these challenges in optimization through the lens of *machine learning*. The resulting approach – *data-driven algorithm design* – has produced effective methods for learning in both exact solvers [1, 2] and heuristic algorithms [3]. While most of my work has been in *machine learning for discrete optimization*, I have also worked on *discrete optimization for deep learning* [4], a promising avenue for research at the intersection of these two fields, with applications in security and resource-efficient machine learning. Simultaneously, I have contributed to the development of machine learning and optimization methods for complex networks, with concrete applications in autonomous vehicles [5], financial systems [6], information propagation [7, 8] and spatio-temporal data [9, 10].

**Research goal.** More broadly, the overarching theme of my research is the tight integration between Artificial Intelligence (AI) and Operations Research (OR), with a special interest in tackling decision-making and learning problems in emerging domains, such as autonomous vehicles and computational sustainability. Research in either direction – AI for OR, OR for AI – requires extensive expertise in both domains, a skill set that I have developed, applied and aim to grow and disseminate in the coming years.

## Research Contributions

Discrete optimization solvers suffer from two main limitations that make them ill-suited for new problems arising in emerging domains, such as those mentioned earlier. On the one hand, algorithmic decisions within the solver are often implemented as handcrafted rules, the design of which requires trial-and-error on a limited benchmark problem set; the set may not include instances that are representative of the user's application. On the other hand, both exact discrete solvers and heuristic algorithms solve each new problem instance *de novo*, even when they have already encountered many similar instances arising from the same application domain.

In [1, 2], I have shown how state-of-the-art mixed integer programming (MIP) solvers can benefit from tailored, efficient machine learning models, resulting in *data-driven* MIP branch-and-bound algorithms that are the first of their kind. The methods I have developed augment the solver with the ability to direct the search based on the characteristics of a particular problem instance and the state of the search, resulting in substantial speedups on a variety of computational sustainability problems, as well as common benchmarks.

**Learning to Branch.** Branching variable selection is a key component of MIP solvers. Choosing the right variables to branch on often leads to a dramatic reduction in the number of nodes needed to solve an instance. An ideal branching strategy (1) gives small search trees, and (2) maintains a low computation footprint.

My work [1] was the first to present an effective machine learning approach to branching in MIP. Given an instance, a variable ranking model is learned on-the-fly, *during the early stages of search*, and is applied thereafter to select a good variable. The learned model essentially approximates "strong branching", a very accurate but time-consuming ranking metric, with a cheap surrogate, thus simultaneously addressing the desiderata above. Most notably, while existing strategies simply score variables based on static, fixed metrics, the learned branching strategy is *adaptive* to the structure of the instance. When used within CPLEX, a commercial MIP solver, the learned branching strategy dramatically reduces the search tree size compared to the widely-used pseudocost branching strategy on a heterogeneous benchmark set (MIPLIB2010). When applied to MIP instances from wildlife conservation planning and road infrastructure design, the learned branching strategy reduces the mean optimality gap from 12% and 15% to 1%, respectively, compared to pseudocost branching.

**Learning to Run Heuristics.** While proving optimality is a key trait of exact solvers, finding high-quality feasible solutions early in the search is at least as crucial. For that reason, MIP solvers use "primal heuristics" periodically during the search. However, the questions of *when* and *what* heuristics should be run during the search are handled heuristically via hard-coded rules: for instance, some heuristics are turned off by default, some run at every node and others every 10 nodes. Such rigid rules are static, instance-oblivious, context-independent, and are unable to adapt to the state of the search.

Alternatively, a heuristic should be run when it is most likely to succeed, based on the problem instance's characteristics and the state of the search. In [2], I study the problem of deciding at which node a heuristic should be run, such that the overall (primal) performance of the solver is optimized. This is the first work that formalizes and systematically addresses this problem. I devised a theoretical framework for analyzing this decision-making question, proposed a machine learning approach for modeling heuristic success likelihood, and designed practical rules that leverage the ML models to dynamically decide whether to run a heuristic at each node of the search tree. This approach improves the primal performance of the SCIP solver by up to 6% on a set of heterogeneous benchmark instances. On synthetic instances of the "forest harvesting problem" from sustainability (framed as generalized independent set) the primal performance improves by up to 60%. Interestingly, these substantial improvements did not require designing new heuristics, but only intelligent, data-driven utilization of existing ones.

When optimality guarantees are not required, practitioners often resort to *heuristics* instead of exact solvers. In [3], I show how a heuristic can be tailored to a distribution of problem instances.

**Learning Greedy Heuristics.** The design of good heuristics for discrete optimization problems often requires significant specialized knowledge and trial-and-error. In many real-world applications, the same optimization problem is solved repeatedly on a regular basis, maintaining the same problem structure but differing in the data (e.g. objective function coefficients). This provides an opportunity for learning heuristic algorithms that exploit the structure of such recurring problems.

In [3], I propose a unique combination of reinforcement learning and graph embedding that address this challenge. Rather than using a simple, static greedy rule to construct a solution to a graph optimization problem (e.g. a greedy insertion heuristic for the TSP), a *learned scoring function* is used instead. Reinforcement learning overcomes the limitations of collecting training data, which requires solving an NP-Hard problem, while the graph embedding approach produces powerful node features that do not rely on feature engineering. The proposed framework can be applied to a diverse range of combinatorial optimization problems over graphs, such as the Minimum Vertex Cover, Maximum Cut and Traveling Salesman problems. The learned heuristics mostly dominate classical algorithms for these problems on a variety of graph distributions, and produce interesting solving behavior that is not typical of classical algorithms.

The work I have presented thus far leverages machine learning in discrete optimization. But can discrete optimization benefit machine learning? I have taken special interest in *discrete neural networks*, a popular class of ML models which requires rethinking vanilla gradient-based optimization methods.

**Fooling Neural Networks with Discrete Optimization.** Recently, it has been shown that neural networks may be overly sensitive to tiny adversarial changes in the input, or "attacks". This weakness is detrimental to the use of these highly-accurate models in safety-critical domains. Designing attack algorithms that effectively fool trained models is a key step towards learning *robust neural networks.* I have designed a novel algorithm for attacking Binarized Neural Networks (BNNs), a class of *discrete deep neural networks* with binary parameters and threshold activation functions. BNNs are popular due to their computational efficiency and potential for deployment onto low-power devices. Attacking BNNs, and consequently protecting them, is thus an important problem in the emerging area of "adversarial machine learning".

The discrete, non-differentiable nature of BNNs, which distinguishes them from their full-precision counterparts, poses a challenge to standard gradient-based attacks. I have studied the problem of attacking a BNN through the lens of combinatorial and integer optimization. First, an integer programming formulation of the problem is derived. While exact and flexible, the MIP quickly becomes intractable as the neural network grows deeper or wider. To address this issue, I designed a decomposition-based algorithm that solves a sequence of small MIP problems, thus scaling much better than the single global MIP. The proposed algorithm vastly outperforms the standard gradient-based attack (FGSM) on a variety of image classification datasets, while simultaneously scaling far beyond a commercial solver on the global MIP.

## Research Agenda

In addressing current and future societal needs, both the public and private sectors are deploying increasingly complex systems at unprecedented scales. Algorithms underlying such systems must evolve and improve rapidly to keep up with the pace. For that, I am excited to continue bringing AI and OR together. On the technical front, my first long-term goal is to devise ML methods that streamline the process of algorithm design for optimization, particularly in new, uncharted domains where classical paradigms may not be effective. Towards that goal, I plan to build solvers that learn continually when deployed in a given environment; create a suite of neural network architectures that can learn flexible heuristics for discrete optimization problems of the future; and design machine learning methods that extract combinatorial structure out of the large amount of instances being generated in high-velocity applications. My second long-term goal is to bring constraint reasoning closer to deep learning, addressing challenging problems such as: training resource-efficient neural networks with integer parameters; simplifying over-parameterized models via pruning; in adversarial machine learning, attacking and protecting neural network models with structured inputs (e.g. sentences in neural machine translation).

**Lifelong Learning Solvers.** Following my work on learning within exact MIP solvers, multiple research groups have developed ML methods for other tasks, such as selecting a decomposition in a Dantzig-Wolfe-based solver, or selecting a good formulation for a non-linear integer program. The bulk of the work in this area consists of an offline model training phase, followed by testing. While it is fruitful to study offline learning approaches for each solver component in isolation, I envision that the full potential of machine learning in improving exact solvers is yet to be fulfilled.

In this line of research, I will build discrete optimization solvers that learn *continually* over the lifetime of their deployment in an industrial domain, i.e. lifelong learning solvers (L$^3$Ss). Instead of adapting a single component of the solver at a time (e.g. branching or heuristics), multiple key components will be automatically tuned to the distribution of instances being encountered. This effort will build on my thesis research, extending the reinforcement learning approach developed in [3] to branch-and-bound decisions. A L$^3$S will use performance metrics such as running time as a signal for improving the reinforcement learning policy that is guiding the search. Practically, a L$^3$S will first be implemented from scratch for prototyping purposes, before integrating the novel learning techniques that will be developed into a full-fledged open-source solver, such as SCIP, which can then support real-world applications in the field.

**Learning Heuristics for Diverse Combinatorial Structures.** Since my paper [3] has appeared, there has been a number of follow-ups that apply similar methods to other problems (e.g. TSP variants such as the Vehicle Routing Problem), or use other neural network architectures. Despite promising results, existing learning methods are only capable of handling simple constraints for very structured problems, e.g. subtour constraints for insertion heuristics in TSP. In practice, however, real discrete optimization models typically involve a mix of constraints of different types (e.g. knapsack and coverage constraints).

I plan to design neural network architectures that learn heuristics for discrete optimization problems with generic constraints. The key challenge here is to embed classical tools from optimization (e.g. projection, rounding) into the neural network to maintain solution feasibility. This very idea fits well within my research philosophy for data-driven algorithm design: instead of trying to learn algorithms from scratch, one should leverage well-understood techniques from optimization to constrain the space of algorithms being learned over. This long-term project has the potential to extend the impact of machine learning to optimization problems of the future, where traditional algorithm design paradigms may not be able to produce an effective heuristic quickly enough.

**Discovering Hidden Combinatorial Structure.** There is empirical evidence for the existence of "backdoors", small subsets of variables whose fixing resolves the rest of the variables in an integer program or a satisfiability instance. These intriguing objects can be seen as a succinct representation of the combinatorial structure of an instance. More practically, a backdoor can be used to guide branching in exact branch-and-bound, or to generate feasible solutions as a heuristic. As expected, finding a backdoor to an instance is at least as difficult as solving the instance itself. This has confined research on backdoors to simple problems or exploratory analyses. Designing efficient procedures for backdoor-finding has thus remained elusive.

I will investigate the problem of backdoor-finding through the lens of deep learning. When many similar instances are available, a sufficiently expressive neural network model can help identify variables that are likely to be in a backdoor. The key challenge lies in designing graph neural network architectures that capture the dependency structure between variables. The aim is to accelerate the discovery of hidden structures in less well-studied optimization problems from emerging applications.

**Beyond the Gradient: Discrete Methods for Deep Learning.** First-order optimization methods such as Stochastic Gradient Descent (SGD) are largely synonymous with deep learning. Indeed, in many settings, SGD and its variants are both practically, and perhaps theoretically, appropriate. There are, however, other emerging settings, where gradients are not available. I have already studied one such setting in [4], and showed that discrete optimization methods are much more effective and natural. There are other challenging problems in this space that I would like to study.

In the context of adversarial machine learning, I am interested in the problem of attacking and "robustifying" deep models with discrete inputs and outputs. For instance, in machine translation or speech recognition, one can consider attacks that modify a *word* slightly in order to cause a large change in the model's predictions. This is an example of a *discrete decision* that may benefit from discrete optimization methods. Another setting involves the training of discrete neural networks, such as the BNNs I studied in [4]. Current training methods approximate the non-differentiable function represented by the network so that SGD (forcefully) works. Instead, I have begun by devising *local search* algorithms for training BNNs, and have obtained promising initial results. Over the longer term, my goal in this research thread is to design scalable, principled methods for simplifying and securing machine learning models.

# References


[1] Elias B. Khalil, Pierre Le Bodic, Le Song, George Nemhauser, and Bistra Dilkina. Learning to branch in mixed integer programming. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016.

[2] Elias B. Khalil, Bistra Dilkina, George Nemhauser, Shabbir Ahmed, and Yufen Shao. Learning to run heuristics in tree search. In *26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[3] Elias B. Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[4] Elias B. Khalil, Amrita Gupta, and Bistra Dilkina. Combinatorial attacks on binarized neural networks. In *submission to ICLR 2019, arXiv:1810.03538 [cs.LG]*, 2018.

[5] Wenwen Zhang, Subhrajit Guhathakurta, and Elias B. Khalil. The impact of private autonomous vehicles on vehicle ownership and unoccupied vmt generation. *Transportation Research Part C: Emerging Technologies*, 90:156–165, 2018.

[6] Acar Tamersoy, Elias B. Khalil, Bo Xie, Stephen Lenkey, Bryan Routledge, Duen Horng Chau, and Shamkant Navathe. Large scale insider trading analysis: Patterns and discoveries. In *Social Network Analysis and Mining*. Springer, 2014.

[7] Elias B. Khalil, Bistra Dilkina, and Le Song. Scalable diffusion-aware optimization of network topology. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.

[8] Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias B. Khalil, Shuang Li, Le Song, and Hongyuan Zha. Fake news mitigation via point process based intervention. *International Conference on Machine Learning (ICML)*, 2017.

[9] Ardavan Afshar, Joyce C. Ho, Bistra Dilkina, Ioakeim Perros, Elias B. Khalil, Li Xiong, and Vaidy Sunderam. Cp-ortho: An orthogonal tensor factorization framework for spatio-temporal data. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2017.

[10] Sucheta Soundarajan, Acar Tamersoy, Elias B. Khalil, Tina Eliassi-Rad, Duen Horng Chau, Brian Gallagher, and Kevin Roundy. Generating graph snapshots from streaming edge data. In *International World Wide Web Conference (WWW)*, 2016.

# Teaching Statement

Elias B. Khalil

Over the past five years, my experiences as teaching assistant, lecturer and mentor have shaped my teaching philosophy. I am a strong advocate for *integrative* teaching in computational disciplines: motivating computational problems through high-impact applications, emphasizing solid theoretical understanding, and encouraging students to design and implement methods or models and test them on interesting problems. During my PhD, I helped design a large graduate course, presented my work to diverse audiences tens of times, and mentored younger students in their first research experiences. I am excited – and prepared – to teach undergraduate and graduate courses on optimization, machine learning and algorithms. I am also looking forward to creating new courses on emerging topics at the intersection of learning and optimization. In what follows, I describe some experiences that have shaped my teaching philosophy, as well as my vision for what and how I plan to teach.

**Teaching Assistant.** I have played an instrumental role in delivering a graduate algorithms course (CSE6140, Fall 2014). As lead teaching assistant (TA), I contributed substantially to a full revamp of the course, as the instructor was teaching it for first time. This included syllabus design, assignment and exam preparation, multiple lectures, and managing a team of 4 TAs. The course had roughly 150 students with diverse backgrounds: on-campus and distance learners; Ph.D., Masters and a few advanced undergraduates; computing and non-computing students. The size and diversity of my CSE6140 class were a welcome challenge: I quickly adopted an *integrative* teaching strategy that aims to connect algorithmic techniques and problems to relatable notions in the real world, which was highly effective with non-computing students. I designed a project on the Traveling Salesman Problem (TSP), where students implemented TSP algorithms (exact, approximation and heuristic algorithms), and evaluated them on the TSPLIB benchmark; the same project has been used twice since, and has been extended to other NP-Hard problems. Currently, I am serving as TA for the same course, and am helping the instructor, who is also teaching CSE6140 for the first time, tune and deliver the materials I helped create in 2014. In parallel to my TA service, I received formal TA training in two courses at the Center for Teaching and Learning at Georgia Tech, where I learned about and practiced various teaching styles and applied them in my courses.

**Lectures and Presentations.** Throughout my Ph.D., I presented my research at over 25 conferences and seminars, and have continually honed my public speaking skills. I have also taken special interest in *visual communication* methods, with the aim of effectively presenting technical ideas to a broad audience. My presentation skills have been recognized with the first prize at the INFORMS 2017 Annual Meeting poster competition [1], the best poster award at the NemFest Workshop on optimization [2], an outstanding poster award twice at the Georgia Tech "Career, Research and Innovation Conference", and the best paper award at a NIPS workshop [3]. I have also served as lecturer on a number of occasions. As a TA, I have given multiple lectures on approximation and submodular optimization. In June 2018, I was invited to give a "Master Class" (tutorial) on deep learning for combinatorial optimization at the 15th International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR). Out of six tutorials, I was the only student presenter to address the 200 conference attendees for 90 minutes.

Through these experiences, I have learned to tailor my presentations appropriately, depending on the audience (students in a classroom or colleagues at a conference), and the nature of the material I am teaching (fundamentals or recent research developments). I believe this flexibility will be key to my delivering the teaching material effectively.

**Courses I'd Like to Teach.** At the *undergraduate* level, I am well-positioned to teach introductory courses in optimization, artificial intelligence and machine learning, as well as data structures and algorithms. At the *graduate* level, I would like to teach advanced courses in *machine learning* and *computational methods in operations research*, particularly integer and linear programming.

I am keen on designing new graduate-level courses on growing research areas within my expertise. Next, I will list some topics that I would like to cover in the proposed courses as a form of mini-syllabus.

- *Integer Optimization for Machine Learning*: this course introduces the use of discrete optimization in machine learning and statistics, both for classical and modern tasks, such as:

[1] http://ekhalil.com/pdfs/poster_informs2017.pdf
[2] https://pwp.gatech.edu/nem-fest-2017/
[3] http://snap.stanford.edu/networks2013/

  - Best feature subset selection
  - Decision tree learning
  - Graphical model inference
  - Natural Language Understanding
  - Verification of deep neural networks

- *Machine Learning for Optimization*: this course touches on learning as applied to continuous and discrete optimization, exact and heuristic solvers. Topics include:
  - Parameter tuning of solvers
  - Iterative algorithms as policies in reinforcement learning
  - Learning in gradient descent
  - Learning in combinatorial optimization
  - Learning in exact solvers

- *Modern Integer Programming Solvers*: this course targets Ph.D. students in operations research, and uses the open-source solver SCIP to investigate the following:
  - Preprocessing and presolving techniques
  - Branching strategies
  - Primal heuristics
  - Cutting plane management
  - Advanced topics: Spatial branch-and-bound for non-linear problems; Parallel branch-and-bound

In both undergraduate and graduate courses, I will use programming assignments and projects to: (1) materialize abstract algorithmic concepts and models taught in class, and (2) prepare students for advanced academic and industrial research in OR and ML. Given my experience with AMPL, CPLEX, Gurobi and other solvers, their interfaces to modern programming languages such as Python, and modern ML toolkits such as PyTorch and scikit-learn, I am very comfortable designing such projects and guiding students throughout.

**Student Advising and Mentoring.** As part of the Georgia Tech CSE CRUISE program in 2015 [4], I advised Mr. Sachin Grover, then a computer science senior at IIT Jodhpur. I guided Sachin through the process of learning about branch-and-bound for integer programming, far beyond his background in optimization and algorithms at the time, resulting in fast turnaround despite the 10-week internship. Perhaps most importantly, I addressed Sachin's questions about graduate school in a foreign country, and motivated him to pursue a research career; he is now a Ph.D. student in computer science at Carnegie Mellon University. Earlier, during my masters studies, I mentored Mr. Samuel Clarke for his first undergraduate research experience at Georgia Tech and introduced him to best practices for conducting research; Samuel went on to receive the prestigious "Astronaut Scholarship" (2015), and is now a masters student at the Robotics Institute, Carnegie Mellon University.

**Beyond the Classroom.** Between 2016 and 2018, I was elected to serve as vice president (VP) of my department's graduate student association (GSA). I was the main organizer of the department's biweekly student seminar, HotCSE [5] (Hot Topics in CSE), which featured 40-minute talks by Ph.D. students in CSE at Georgia Tech, for a total of 25 talks during my tenure. I actively recruited first and second-year Ph.D. students to present at the seminar, so as to bolster their confidence and provide them with the opportunity to practice their public speaking skills, particularly before their first conference talks. Often times, the seminar featured talks of different nature (theoretical or computational) or focus (e.g machine learning or computational biology), so as to attract students with diverse interests and expose them to their colleagues' research areas.

**Conclusion.** In summary, I am excited about the opportunity to teach students about optimization, algorithms and machine learning. My approach to teaching will be inclusive of the students' diverse backgrounds and levels, with the aim of nurturing their passion for science and engineering, and developing their capabilities for rigorous problem-solving in the face of pressing societal and environmental challenges.

[4] https://www.ic.gatech.edu/news/432091/international-students-cruise-through-summer-georgia-tech
[5] http://www.hotcse.gatech.edu/