ORMS TOMORROW

informs Student Magazine

Machine Learning and Operations Research

- ✓ An opportunity for new directions and to engage with new areas
- ✓ ML and mixed integer programming
- ✓ Is ML really the hero that the OR community needs?
- ✓ Data-centric versus problem-centric methods
- ✓ ML applications in the energy sector
- ✓ ML applications in material science
- ✓ ML and ethnography: a marriage made in heaven
- ✓ INFORMS Student Chapter spotlight: Northeastern University
- ✓ INFORMS Student Chapter highlights

@ORMS_Tomorrow
facebook/ORMSTomorrow
linkedin/groups/3788040

Spring/Summer 2018

ORMS TOMORROW informs Student Magazine

Spring/Summer 2018

Dear Readers,

We are pleased to release the Spring/Summer 2018 edition of OR/MS Tomorrow. The creativity and hard work from the entire staff at OR/MS Tomorrow has made this possible. This issue focuses on machine learning (ML) and operations research (OR). The past few years has seen a proliferation of work in the exciting intersection of ML and OR, both in terms of methodological research and practical applications. This dedicated issue highlights the use of ML in specific domains such as in the energy sector, material science, ethnography, etc., and discusses recent work in integer programming and ML, in addition to the perception of ML from the OR side. We are excited for you to read this issue. Over the past year, we have not only grown in terms of our staff size, but have also expanded our online presence through consistent social media efforts. We are excited to be providing you with the best perspective of news from the realm of Operations Research and Management Sciences. We hope you enjoy the content and look forward to hearing any thoughts you may have regarding OR/MS Tomorrow via email at `orms_tomorrow@mail.informs.org`.

Hossein Badri and Rahul Swamy
Co-Lead Editors

Machine Learning – An Opportunity for New Directions and to Engage with New Areas

Emily Tucker
Ph.D. Candidate,
Industrial and Operations Engineering,
University of Michigan

MACHINE learning (ML) is part-buzzword, part-powerful algorithmic toolbox. It's hyped to change the world, and people in many fields are beginning to use its techniques to understand their systems and make better decisions. As operations researchers, we spend our careers developing models to provide insights. There is a range of overlap between ML and operations research (OR), and I think the popularity of machine learning gives us the opportunity to work in new areas and solve broader, more complex problems.

One of my friends who is a biologist once told me about a conversation she had with her PhD advisor; while they were discussing possible methods, her advisor at one point responded with a wave of her hand, "Oh, why don't you just do some machine learning on that." But neither knew much of the specifics!

Even a passing mention can be an open door to a discussion of new techniques. While folks may not have heard of operations research, we can discuss how "algorithms" can be used in other ways as people become more comfortable with the term. This increase in familiarity may help us work with our clients and collaborators to choose the best methods for their problem at hand, whether it involves clustering, or integer programming, or other techniques.

Many fields are beginning to use machine learning to solve complex problems. Astrophysicists are working to classify galaxies, and information scientists are analyzing literary text ("How Machine Vision Is Reinventing the Study of Galaxies," 2015; Prospero, 2018). The Seattle Seahawks, an American football team, use ML to try to prevent injuries (Soper, 2017). Researchers are studying defensive strategies in the NBA ("Machine Learning Proves Useful for Analyzing NBA Ball Screen Defense," 2016). Recently, Nature published a paper that discussed how supervised machine learning and matching algorithms had improved the refugee assignment process (Bansak et al., 2018).

We don't need to go far afield though to find new directions as machine learning naturally parallels many of the traditional applications of operations research. Businesses have long struggled to manage "customer churn" by retaining and recruiting new customers, and ML algorithms can enhance traditional models by incorporating a wide range of data to understand customer dynamics, including analyzing click rates and detailed order histories.

These insights can improve decision-making in perhaps unexpected ways (Neff, 2014). Walmart has found that certain weather conditions correlate with food purchases; their steaks tend to sell when it is warm and windy whereas hamburgers do better when it is warm but less windy. In turn, Walmart can tailor its advertising down to a zip code-level based on weather conditions, and as a result, sales have gone up.

There have also been mountains of OR work in inventory management, and researchers are starting to use ML and deep learning techniques to improve order quantities and timing (Snyder, 2018). Rather than separating the problems of estimating demand distributions and optimizing operational decisions, researchers have integrated the two by incorporating ML techniques and found that they can substantially reduce cost.

ML is popular partly because of its impact and partly because it is relatively easy to implement, particularly in contrast to many optimization algorithms. Within R or Python, if you download the appropriate package, only a few lines of code are needed to start using machine learning.

A word of caution as we dive deeper, however the relative ease with which ML algorithms can be applied can obscure potential biases in their insights. If we're not careful, these biases can have major repercussions on perpetuating inequality (Mok, 2017) and may cause more problems than they fix!

If you are interested in learning more, several online resources including popular courses from Coursera (Ng, 2018) and helpful notes from Chris Albon (Albon, 2018) provide instruction in ML techniques. Many INFORMS student chapters are also organizing sessions and workshops on machine learning. Check out the chapter highlights in this edition for a few examples.

Machine learning and operations research are natural neighbors (one might even argue nearest-neighbors?), and I believe there's an important and unique role for both. Even basic familiarity with machine learning may open the door for OR professionals to start new and unexpected collaborations.

References

1. Albon, C. (2018). Notes On Using Data Science & Artificial Intelligence To Fight For Something That Matters. Retrieved May 15, 2018, from https://chrisalbon.com/#machine_learning

2. Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. Science, 359(6373), 325–329. https://doi.org/10.1126/science.aao4408

3. How Machine Vision Is Reinventing the Study of Galaxies. (2015). Retrieved May 15, 2018, from https://www.technologyreview.com/s/536411/how-machine-vision-is-reinventing-the-study-of-galaxies/ Machine Learning Proves Useful for Analyzing NBA Ball

4. Screen Defense. (2016). Retrieved May 15, 2018, from http://www.eecs.umich.edu/eecs/about/ articles/ 2016/ MIT-Sloan-Sports-Analytics.html

5. Mok, K. (2017). Mathwashing: How Algorithms Can Hide Gender and Racial Biases. Retrieved May 15, 2018, from https://thenewstack.io/hidden-gender-racial-biases-algorithms-can-big-deal/

6. Neff, J. (2014). Cloudy With a Chance of Meatballs: How Weather Forecast Predicts Walmart's Sales Outlook. Retrieved May 15, 2018, from http://adage.com/article/dataworks/ weather-forecast-predicts-sales-outlook- walmart/295544/

7. Ng, A. (2018). Machine Learning. Retrieved May 15, 2018, from https://www.coursera.org/learn/ machine-learning

8. Prospero. (2018). Machines are getting better at literary analysis. Retrieved May 15, 2018, from https://www.economist.com/prospero/2018/03/08/machines-are-getting-better-at-literary-analysis

9. Snyder, L. (2018). Machine Learning for Inventory Optimization. Retrieved May 15, 2018, from http:// opexanalytics.com/ machine-learning-inventory-optimization/

10. Soper, T. (2017). How the Seahawks use Microsoft's new high-tech performance platform to prevent injury and plan practices. Retrieved May 15, 2018, from https:// www.geekwire.com/2017/seahawks-use-microsofts- new-high-tech-performance-platform-prevent-injury- plan- practices/

Machine Learning and Mixed Integer Programming

Rahul Swamy
Ph.D. Candidate in Industrial Engineering,
University of Illinois at Urbana-Champaign

In today's world, advancements in computing power augmented by the availability of data has ushered in an unprecedented ability to transform data into useful insights. Machine learning (ML) has earned its place as a quintessential tool in any data scientist's toolbox. Mixed Integer Programming (MIP) has provided a long-standing framework for solving large NP-Hard problems to theoretically-proven optimality and has revolutionized many industries such as logistics and transportation. Even though both ML and MIP share a common trait - using data to influence decision making - they have [for the large part] been studied by different research communities. Squarely in the interface between computer science and operations research, interesting problems have emerged when studying ML and MIP together.

The idea behind ML had existed since 1950s but found prominence much later with the merging of statistical tools, continuous optimization and big data. With the boom in Artificial Intelligence, present day ML has transformed into variations of Deep Neural Networks (DNNs) and has found success in applications such as in personal assistants, product recommendations, spam filtering, among others. For MIPs, the classical branch-and-bound technique has seen remarkable speed-ups over the last half a century with theoretical breakthroughs in cutting plane theory, decomposition techniques, column generation, among others, in tandem with efficient implementations in commercial solvers. What were once considered "unsolvable" problems can now be solved in a matter of seconds. For example, the branch-and-cut-based solver Concorde TSP can solve TSP instances with more than 85,950 cities as reported in Applegate et al. (2009). At the outset, ML and MIP may seem to have different objectives. ML predicts while MIP solves. However, there are fundamental problems in both the fields that cross-disciplinary research can be benefit each other.

Continuous optimization forms the crux of (supervised) ML, with the training phase using a fraction of the data to learn (optimize) for certain model parameters. However, the use of discrete optimization and its solution strategies has been relatively underexplored in ML literature. Misquoting a visionary who once shot for the moon, "Ask not what ML can do for MIP - ask what MIP can do for ML." In that spirit, some of the MIP for ML works are highlighted here.

Support vector machines (SVMs) are popular classification techniques in ML. Early work by Bennett and Demiriz (1999) posed semi-supervised SVM as an Integer Program and solved it using CPLEX. For the classic SVM problem with ramp-loss minimization, Belotti et al. (2016) present a reformulation to tackle the big-M constraints. Üney and Türkay (2006) provide an MIP approach for a general multi-class classification problem using hyper boxes to partition the data. Word alignment is a key component of language translation tools. Lacoste-Julien et al. (2006) solve this problem by framing it as a Quadratic Assignment problem. DNNs have gained popularity in recent years. As a step towards modeling DNNs using MIP, Fischetti and Jo (2018) present an MILP model with application to feature visualization and adversarial ML. The latter is a natural setting since it needs optimal solutions that "fools" the DNN by overfitting it.

On the other hand, if one indeed asked what ML can do for MIPs, there are answers to that as well. The conventional branch-and-bound procedure involves a host of parameters that are typically tuned ad-hoc to the needs of specific [classes of] problems.

These parameters decide what the next step should be at a certain node in the branch-and-bound tree. Recently, several papers explored how ML can be used to tune these parameters that decide different aspects of branching and bounding - Kruber et al. (2017) on deciding whether Dantzig-Wolfe decomposition should be used or not, Khalil et al. (2016) on which variable to branch on, Alvarez (2016) on branching decisions specific to approximations to strong branching and Khalil et al. (2017) on whether a primal heuristic should be used or not. The idea here is to predict whether making a decision at a node will improve the overall run-time of the algorithm. The results from these works provide promising directions for not only generic MIP solvers, but also for exploring their success in specific classes of problems.

At his recent plenary talk at EWGLA 2018, Prof. Andreas Lodi from Polytechnique Montreal posed an interesting perspective that ML and optimization are "two-sides of the same coin". There is a dual nature to an algorithm that solves an optimization problem and the parameters that can make it efficient, and there is a need for frameworks that integrate ML and optimization. While this article is nowhere close to a literature review, the goal is to touch upon some of the opportunities in the intersection of ML and MIP. There are many more interesting open challenges that must be of interest to both communities and the future will tell.

References

1. Alvarez, A.M., Wehenkel, L., and Louveaux, Q. (2016). "Online learning for strong branching approximation in branch-and-bound."

2. Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W., Espinoza, D.G., Goycoolea, M., and Helsgaun, K. (2009). "Certification of an optimal TSP tour through 85,900 cities." Operations Research Letters 37.1: 11-15.

3. Belotti, P., Bonami, P., Fischetti, M., Lodi, A., Monaci, M., Nogales-Gómez, A. and Salvagnin, D. (2016). "On handling indicator constraints in mixed integer programming." Computational Optimization and Applications 65.3: 545-566.

4. Bennett, K.P. and Demiriz, A. (1999). "Semi-supervised support vector machines." Advances in Neural Information processing systems.

5. Lacoste-Julien, S., Taskar, B., Klein, D. and Jordan, M.I. (2006). "Word alignment via quadratic assignment." Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics.

6. Khalil, E.B., Bodic, P.L., Song, L., Nemhauser, G. and Dilkina, B. (2016)."Learning to Branch in Mixed Integer Programming." AAAI.

7. Khalil, E.B., Dilkina, B., Nemhauser, G., Ahmed, S. and Shao, Y. (2017). "Learning to run heuristics in tree search." Proceedings of the international joint conference on artificial intelligence. AAAI Press, Melbourne, Australia.

8. Kruber, M., Lübbecke, M.E. and Parmentier, A. (2017). "Learning when to use a decomposition." International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. Springer, Cham.

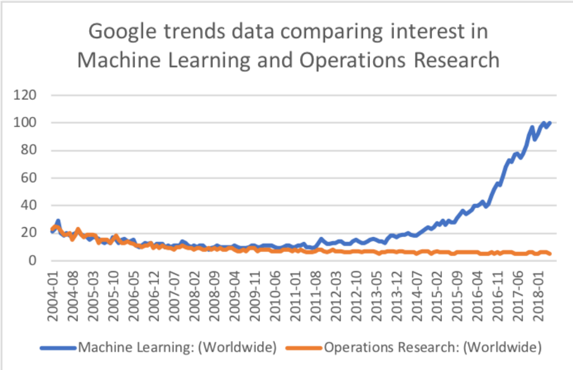
9. Üney, F. and Türkay, M. (2006). "A mixed-integer programming approach to multi-class data classification problem." European journal of operational research 173.3: 910-920.

Machine Learning: Is it really the hero that the Operations Research community needs?

Siddhartha Nambiar
Ph.D. candidate in Industrial Engineering,
North Carolina State University

About a year ago, I was presenting a poster outlining my current research at a conference. My work was about an application of Markov Decision Processes, and I had had several people stop by at my poster to interact with me. Most of the questions that came my way were easy to deal with, until one individual asked me something that motivated me to write this article. His question was simple – why not just use machine learning to do this? I'm not going to tell you how I responded to him because that would require I outline parts of my research which is outside the scope of this article. What I will tell you is that it got me thinking about how Machine Learning as a buzzword is used in so many different contexts today. A fellow student once mentioned to me that he felt graduate students in Operations Research (OR) and Statistics are at a substantial handicap when compared to graduate students in Machine Learning (ML), despite being in substantially overlapping subjects. While I wasn't sure if I agreed with the first part of his statement, the second part resonated with me. Both ML and OR share foundations in probability theory, optimization, and linear algebra. There is little doubt that both OR and ML possess their own independent domains. OR, as the application of scientific and mathematical methods to the analysis of complex systems, was invented during World War II. ML, as the branch of computer science that gives computers the ability to learn without being explicitly programmed, is more recent and has been around since the 80s.

Since its inception, however, ML has been slow to catch on. Comparing the interest between the two fields shows ML only break away after 2010. Perhaps advances in computational capabilities played an important role in the rise in popularity of Machine Learning. The interesting question then becomes – why has this not extended to Operations Research? The perceptions among practitioners about what the two topics are and how they contribute to the research community are variable. As one researcher at Google explains, perhaps the difference between the two fields is just semantics. His point is that, at the end of the day, people are people and could as easily work under the umbrella of OR as ML. However, he admits that some ML algorithms sometimes tend to be 'hacks'; they rely on a ton of methods that depend on intuition rather than theory.



A paper published in the Journal of Machine Learning Research in 2006 (Bennet and Parrado-Hernandez, 2006) looked at how the two fields are largely intertwined. The authors state that ML researchers have embraced the advances in optimization thus allowing new types of models to be pursued. The natural thought that follows then, is that the difference between the two fields is not as rooted in theory as it is in semantics and accessibility, the latter being an important factor. Today, a very large number of people recognize the term Machine Learning. While this is largely in part due to our world being overrun by tech companies that use the term as a means of catching the public's attention, nuances in the way that ML is being taught cannot be ignored. As it stands today, ML is not something that too many people understand. When I think about the individual who suggested I apply ML to my research at the conference, I can't blame him. The world we live in is more about 'using ML' rather than 'learning ML'. In other words, give the man a fish rather than teach him to fish, because teaching him to fish would involve teaching him multi-dimensional calculus, advanced linear algebra, optimization theory, and advanced probability theory, just to start with. Whether this is a problem or not is up to the community to decide. Proponents of the status quo will argue that this is how academic and technical research has been taking place for many years and that ML is no different. But is ML really 'no different', especially now that policy makers in congress are starting to pay attention to the way ML and AI are shaping the lives of the public?

In summary, Machine Learning is more popular than Operations Research because of two reasons. The first is that the biggest part of the analytic decision-making process has now been partially shifted to the machine. The investment in machine learning is a natural evolution in technology and humanity's demand to create technologies that extend our own capabilities. And while the same can be said of advancements in Operations Research, the second reason for ML's popularity is what really seals the deal – the fact that it works! And not just in the sense that it provides us with excellent analytic solutions to difficult problems, but in the sense that the popularity has caught on. A direct outcome of this meteoric rise in popularity is that more and more research is currently focused on improving existing methods. There is no doubt that OR benefits from this as well, since it has become increasingly coupled with ML. However, whether the semantics of OR can survive the current outburst in ML interest remains to be seen.

References

1. Bennett, K. P., & Parrado-Hernández, E. (2006). The interplay of optimization and machine learning research. Journal of Machine Learning Research, 7(Jul), 1265-1281.

The Use of Quantitative Methods with Two Different Perspectives: Data-Centric versus Problem-Centric

Çağlar Çağlayan
Ph.D. in Operations Research,
Georgia Institute of Technology

DESPITE the diversity of the data analytics methods and the variety of business problems, the quantitative decision science methods can be grouped into two main categories: data-centric and problem-centric approaches (Wegryn, 2014; Rose, 2016). The ultimate purpose of both data- and problem-centric approaches is the same: to help decision-makers make better (informed, effective and efficient) decisions. Yet, usually, the way they approach a problem is fundamentally different (Figure 1). The goal of this article is to briefly introduce these two "quantitative decision science" approaches, highlight their differences, and discuss their roles in better decision-making.

Data-centric approaches, as the name implies, prioritize the use of data and aims to gain insights from the data about the problem of interest. Accordingly, the efforts of the practitioners of data-centric approaches are primarily concentrated on

(1) analyzing (including cleaning, summarizing, and manipulating) data; (2) investigating the extend of its use and limitations; and (3) generating new insights from the dataset by applying (and tailoring) analytical methods, ideally based on justifiable assumptions that can be validated. Some means of translating data into new insights, by applying data-centric approaches, are as follows: Identifying and quantifying relations between a key outcome and variables in the dataset (statistical association), detection of systematic changes that variables exhibit over time (pattern recognition), and estimating the future course of a key outcome by modeling the behavior of certain variables (prediction/forecast). A few examples of data-centric techniques are time-series analysis, regression models, machine learning methods, and deep neural networks. The use of data is also

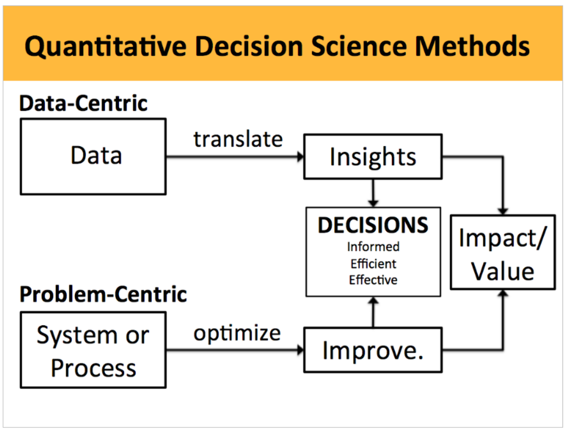


Figure 1: Data- and Problem-Centric Approaches

critical for applying problem-centric approaches to business problems as the qualities of a dataset (e.g., its size, the uncertainty around its variables, etc.) significantly affect the choice of the analytical model to be used. Yet, instead of data, the primary focus of the practitioners of problem-centric approaches is on the business problem itself. The primary goal is to convert the business problem into a well-defined analytical problem that can be modeled and solved. Identifying the decisions to be made, the key outcome to be improved, and understanding the mechanisms governing the key dynamics of the business problem are usually the initial steps of problem-centric approaches. These steps are followed by the development of an analytical model that captures the key dynamics of the business problem and links these dynamics to an objective function to be optimized through the decision variables. The final step is to use an algorithm that solves the analytical problem and identifies the optimal decisions generating the best result(s). A few examples of problem-centric methods are as follows: deterministic optimization (e.g., linear and integer programing), discrete-event simulation, queuing theory, and Markov decision processes.

Understanding the critical features of the business problem and the content and limits of the dataset are required steps, up to a certain extent, both for data- and problem-centric approaches. Whether the primary efforts are concentrated on extracting information from the data or mathematical modeling of the problem is where data- and problem-centric methods begin to differentiate. One way to see the difference between these two approaches is to look at what kind of questions they address through their validation (and debugging) efforts. A typical question for a data-centric approach is the following: Does the employed quantitative method describe the patterns and relationships that the dataset manifests with a high level of accuracy/precision [and hence, can be trusted to make future predictions]? On the other hand, a standard question for a problem-centric technique is as follows: Does the utilized analytical model correctly capture the key dynamics of the underlying problem without over-simplification [and hence, can be used to identify the best course of action]? As it can be seen by these two sample questions, a primary concern regarding the validity of one approach is mainly on the correct use of the data whereas the validity of the other is challenged via its capability to capture of the key problem features. Accordingly, while debugging, a data-centric method might be dealing with over- and under-fitting issues while a problem-centric approach might need to address problems such as a missing constraint or a wrong objective function.

The differences between data- and problem-centric approaches are also related with their roles and objectives. To explain these differences better, we can get assistance from a few terms: descriptive, diagnostic, predictive and prescriptive analysis (Maydon, 2017).

- **Descriptive Analysis:** The quantitative description of important information contained in the dataset.
- **Diagnostic Analysis:** Examination of the historical course of the process of interest and identification the relations of system behavior and process outcomes with the variables in the dataset.
- **Predictive Analysis:** Projection of the future behavior of the process as a function of certain variables.

- **Prescriptive analysis:** Identification of the best course of action to be taken to improve the system/process of interest.

Generally, descriptive and diagnostic analyses are conducted via data-centric methods; prescriptive analysis is performed by problem-centric techniques; and both data-centric and problem-centric methods are utilized for predictive analysis.

To conclude, there is a wide range of variety in analytical methods employed by researchers and practitioners to solve their business problems. Despite the variety, the "quantitative decision science" techniques can still be placed into one of the following categories, based on how they approach a problem: data-centric and problem-centric. Generally, these two perspectives not only have different ways to help decision-makers, but also conduct analyses at different dimensions and hence, have different roles at generating better data-driven decisions. Although it might be unrealistic to expect from any practitioner to be an expert of both approaches, it is very beneficial (if not required) to be familiar with some analytical methods in both domains to correctly approach a business problem and to have a better (full-picture) understanding of the utility of quantitative methods for decision-making.

References

1. Wegryn, G. (2014). Top 5 Analytics Trends for 2015. INFORMS Podcast.
2. Rose, R. (2016). Defining Analytics: a Conceptual Framework. OR/MS Today, 43(3), 36-41.
3. Maydon, T. (2017). 4 Types of Data Analytics. Principa, Url: https://insights. principa.co.za/4-types-of-data-analytics-descriptive-diagnostic-predictive-prescriptive, accessed June 2018.

Machine Learning Applications in the Energy Sector

Sepehr Ramyar
Department of Technology & Information Management,
University of California Santa Cruz

THE energy industry, and particularly the power sector, is undeniably an essential component of any modern society. In fact, electrification has been determined as the "greatest engineering achievement of the 20th century" by the National Academy of Engineering (NAE, 2018). However, this mighty engineering achievement of the previous century faces just as great of challenges in the 21st century that can no longer be addressed using analytical tools of the past. Increasing scale and complexity of the power systems, inclusion of new active players and stake-holders in the power sector, and rapid evolution of institutions and regulations that govern and operate the energy industry are the main reasons that necessitate the application of robust analytical tools that can efficiently and effectively address these challenges. Machine learning has proved to be a powerful tool in extracting and processing information from large sets of data and has been used extensively for different applications in the energy sector. In this article, we will discuss how machine learning has been applied to enable a more efficient power system.

The power system traditionally was a monopoly in which a single utility company produced, transmitted, distributed, and ultimately sold electricity to end-users. This paradigm, however, started to change in the 1980's and 90's as competition was introduced to the power systems to induce economic efficiency. In the new framework, power is traded in an electricity market, usually called the wholesale market, in which generators and consumers (the load serving entities[1]) bid their supply and demand quantities and the market consequently clears under a set of conditions[2], optimally allocating power by merit order: the cheapest producers (generators) supply electricity to the consumers with the highest willingness-to-pay. In technical terms, this is a type of double auction. For example, imagine there are three generators respectively offering 5, 10, and 15 megawatts (MW) for 3, 4, and 5 dollars per MW ($/MW) for a specific hour of the day. There are also three consumers bidding 1, 3, and 6 $/MW each demanding 5 MW for the same time/hour of the day. Then, the least-cost generator, i.e. the 3 $/MW generator, is allocated its full 5 MW to meet the demand of the consumer with the highest willingness-to-pay, i.e. the 6 $/MW consumer. Because the remaining two consumers' bids are lower than the remaining two generators' asks per MW, the generators are not able to further supply the remaining customers. Thus, the market clears at 3 $/MW. Each day, a similar process is repeated for each of the 24 hours of the following day in what is called a day-ahead market to determine the price and quantity of the power to be generated and delivered.

It is evident that if one is able to accurately predict the market clearing price, then they can place bids that have a much higher chance of be-

[1] Load serving entities are electric service providers, i.e. retail electric providers or municipally owned utilities.

[2] This is usually referred to as market clearing conditions that in addition to trying to set demand equal to supply, also involve physical characteristics of the power grid in terms of capacity, congestion, etc.

ing cleared (and compensated). Moreover, the Independent System Operator (ISO) that operates the power market wishes to identify and intercept fraudulent bids to limit strategic behavior by participants. As mentioned before, this auction is carried out every day for each of the 24 hours of the following day. This means that there is plenty of data available to exploit and this is where machine learning tools come in handy. One particularly useful method is Support Vector Machine (SVM). This machine learning technique maps inputs to a feature space and then the predicted outcome is calculated as a linear function in the new feature space. The strength of this method lies in its ability to linearly separate a dataset even though it may not be separable in the original space of the training data. In the case of power price forecasting, the feature space could include elements such as time (day and/or hour), temperature (max, min, average), humidity, etc. in the location of demand. The linearity saves a lot of computation time and makes SVM a particularly useful analytical tool in the context of an electricity market. Specifically, in the real-time market[3] (as opposed to day-ahead markets) the entire bidding and market clearing process is carried out in a very short time window (less than five minutes) and consequently it is essential for market participants to be able to carry out massive computations for predicting the price is the shortest time possible, and this is where SVM's computational efficiency comes in.

Hourly prices for power through a market mechanism introduce a host of opportunities for new businesses to grow in the energy sector. Once particular example is energy demand management. This stems from the fact that there are different energy-consuming activities (e.g. lighting, heating, electric vehicle charging, etc.) and each consumer has his own preferences/ranking over them or can be incentivized to rank them. Using machine learning algorithms, it is possible to recognize the energy consumption pattern of each consumer and identify energy savings opportunities based on the variable market prices. For example, one can automatically, using smart energy management systems that run on machine learning algorithms, shift their electric vehicle charging from peak hours to off-peak hours with lower prices. One particularly useful machine learning tool here is reinforcement learning. This method dynamically learns and adopts the behavior that yields maximum reward which in this context could be translated into monetary savings. The applicability of the reinforcement learning in demand management is that no historical data is required, and the algorithm would be able to navigate and detect optimal action in real-time.

Scaling up this solution, a new participant emerges in the power market: aggregators. An aggregator operates vast fleet of demand management systems throughout a community. Each of these units is capable of identifying consumption patterns and energy saving opportunities based on the machine learning algorithms. This energy saving could be in form of forgoing consumption (e.g. reduced lighting) or shifting consumption (delaying electric vehicle charging to off-peak hours). The aggregator then bundles or aggregates these energy savings and offers it to the wholesale market in form of demand response i.e. power that otherwise had to be generated and consumed and receives compensation in return. In addition, when the power grid becomes under stress, the ISO can issue demand response (ask for reduced demand) and aggregators would then be able to secure the required demand response from their pool of customers and receive (and share) compensation for it. This guarantees reliability and efficiency of power systems.

So, we can see how different elements of the current institutions that operate power systems can be improved using machine learning algorithms. By enabling the computation and extracting information from massive data sets, machine learning algorithms have enabled various agents in the power system such as consumers, generators, market operator, and aggregators to scale up solutions and effectively exploit new opportunities introduced in the new energy sector paradigm while at the same time improving economic efficiency and reliability of the greatest engineering achievement of the 20th century.

References

1. NAE [National Academy of Engineering]. (2018). Greatest Engineering Achievements of the 20th Century. Retrieved from http://www.greatachievements.org/

[3] After the day-ahead market clears and quantities are determined, there might still be deviations from the allocations due to unexpected failures or demand surge. These deviations are settled in the real-time market on the same day that units are dispatched.

Machine Learning in Material Science

Shreya Gupta
PhD Candidate,
Department of Mechanical Engineering,
University of Texas at Austin

A little over a month ago I was connected with Dr. Yuanyue Liu who is an Assistant Professor in the Department of Mechanical Engineering at The University of Texas at Austin. His research focuses on fundamental and technological problems in material science related to electronics, optoelectronics, energy conversion and energy storage (e.g., transistors, solar cells, batteries/supercapacitors, electro/photoelectro-catalysis) as well as emerging materials like 2D materials and topological materials. His most recognized work involves the development of a one-step, scalable approach for producing and patterning porous graphene films with three-dimensional networks from commercial polymer films using a CO2 infrared laser (Lin et al. 2014) and the discovery that oxygen on the Cu surface substantially decreases the graphene nucleation density by passivating Cu surface active sites (Hao et al. 2013). We were connected because Dr. Liu needed some assistance in feature ranking and engineering, i.e. ranking predictor variables by their impact on the output variable (feature ranking) and creating new variables from the existing ones that could better predict the output variable (feature engineering). Dr. Liu shared many papers with me on how machine learning (ML) is being employed in material science, and I was encouraged to dedicate a short article to this new area of application for ML. So here it goes! Almost all the papers I read described empirical testing as very costly and time consuming (Faber et al. 2016; Li et al. 2017) even though it is the reason behind the discovery of all industry catalysts known today (Nørskov et al. 2011). Thus, ML algorithms for predicting molecular properties are being increasingly explored and helping in progressing material science at a faster rate than in the past. For example, Ramprasad et al. (2017) cites pioneering applications of machine learning in prediction band-gap of insulators, classification into sp-block and transition metal elements, models to identify correlations and analytical relationships between the breakdown field and other easily accessible material properties such as the band gap and the phonon cutoff frequency. And the list goes on. Faber et al. (2016) highlights that even first-principles methods such as density functional theory (DFT) for computational prediction of the existence and basic properties of crystals composed of only the main group elements (columns I to VIII in the periodic table) is challenging as just these elements lead to approximately 2 x 106 possible elpasolite crystals[4] that can potentially be made. However, ML models are being developed with accuracies close to those of DFT, and only take milliseconds for computations (Montavon et al. 2013; Rupp et al. 2012). Of course, the datasets representing material properties are also small because it's hard to harness data in this field (Ramprasad et al. 2017). Additionally, ML models overcome the trade-off between the versatility of quantum mechanical models and the relative simplicity of semi-empirical force fields. Quantum mechanical models theoretically can be used to study any material as they are governed by analytical differential equations, but these equations are very complex; in contrast, semi-empirical force fields, based on a combination of experimental data and electronic structure calculations on small molecules (Shell 2012), are several orders of magnitude faster but not as versatile. Thus, semi-empirical force fields do not perform well on materials for which the original parameterization were not developed. As ML models are both fast and transferable (Ramprasad et al. 2017), they are gaining popularity amongst material scientists (Faber et al. 2016).

Domain experts and their ML collaborators are also focusing on feature engineering. There is an increasing emphasis on the billions of linear and non-linear compound descriptions that could be engineered using algebraic combinations and mathematical functions (Ramprasad et al. 2017). This would immediately take us into the space of feature ranking. I saw the least absolute shrinkage and selection operator (LASSO) and kernel ridge regression being used widely. In fact kernel ridge regression was used in many models I read about because it works well when attempting to incorporate non-linear relationships. An important aspect of feature selection and engineering that Ramprasad et al. (2017) talks about is the need for feature invariance to certain transformations (some examples of such transformations are spatial rotation, rigid translations, etc.). One of the models that I found interesting was a kernel ridge regression-based ML model developed by Faber et al. (2016) for modeling "the energy difference between the crystal energy and the sum of static, atom-type dependent, averaged atomic energy contributions, obtained through the fitting of each atomic species in all main group elements up to" Bismuth (Bi). They built and employed ML models of formation energies to investigate all possible elpasolites made up of main-group elements. In their paper they present numerical results for ap-

[4] Elpasolite is the predominant quaternary crystal structure, AlNaK2F6 prototype, reported in the Inorganic Crystal Structure Database.

proximately 2 x 106 formation energies which, as discussed earlier, would certainly have been very challenging using first-principles methods like DFT. Li et al. (2017) talks about exploiting the a priori estimation of chemical reactivity of surface metal atoms given the hierarchical complexities in catalyst design. They build an artificial neural network model (see Figure 2) chemisorption model that captures complex, non-linear adsorbate/substrate interactions and thus facilitates exploration of large number of catalytic materials.

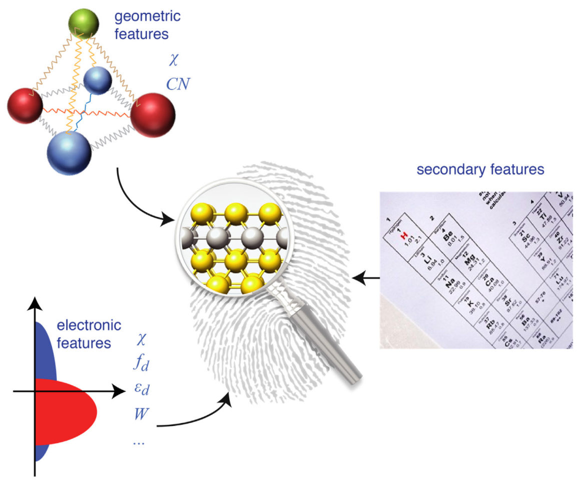


Figure 2: Figure adopted from Li et al. (2017). The authors here are representing a schematic of their neural net "accelerated catalyst design approach."

Sendek et al. (2016) ran all possible LR models for feature selection. They spoke about the difficulty of creating features and eventually with about 20 features they went on to build a logistic regression (LR) model to classify superionic materials based on ionic conductivity. They were careful to include negative examples (i.e., they purposely added many poor conductors to the data used for training and testing) as suggested by Raccuglia et al. (2016). In addition, since they had a very small dataset and a simplistic LR model, they ran LR models with all possible combinations of features and attempted to select the best models using the best LR model. This quickly led to $\sum_{n=1}^{20} \binom{N}{k} = 1,048,575$ models being tested. Finally, they chose the model with the least misclassification rate using metrics such as the training misclassification rate between the predicted and observed the cross-validated misclassification rate using leave-one-out cross-validation (LOOCV).

In conclusion, I noticed that kernel ridge regression and LOOCV were popular modeling approaches (later due to scarcity of data). Ramprasad et al. (2017) provides a survey of many

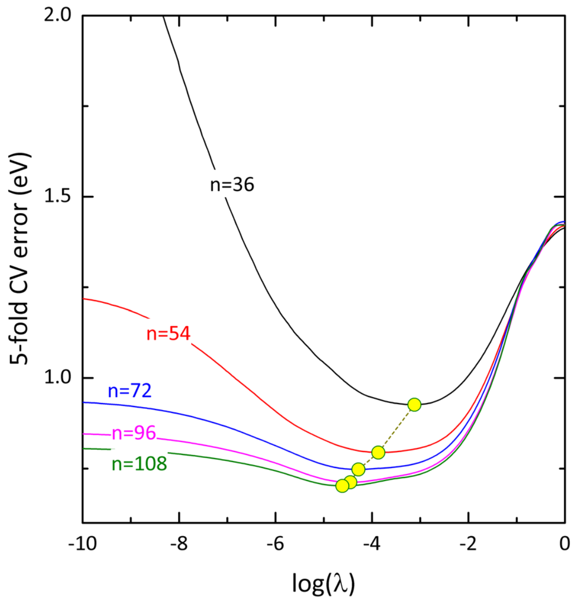


Figure 3: Lasso root mean square error results on 5-fold cross-validation done by Zhang and Ling (2018)

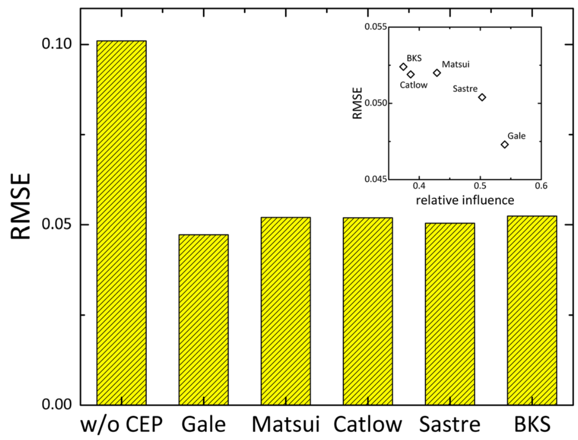


Figure 4: Root mean square errors from a gradient boosted model implemented by Zhang and Ling (2018)

more applications of classification, clustering, regression, etc., in the material science community. They also highlight that future work can focus on building adaptive models that can handle new data points while updating themselves easily but also producing strong predictions for cases where data is different from all prior information (this is the canonical bias-variance trade-off in machine learning). They also talk about the need for uncertainty quantification, the importance of elucidating uncertainty in predictions, and the scope for inverse modeling. Zhang and Ling (2018) also nicely discuss many ways of employing ML in material science where they explore multiple methods like LASSO and gradient boosted trees (see Figures 3 and 4). One of the main hurdles facing the material science community when employing ML is insufficient data (Ramprasad et al. 2017; Zhang and Ling 2018). Generating data is time consuming and expensive, and material science datasets often are very wide. However, there is a lot of scope for building ML models that can help progress research in material science at a faster and cheaper rate.

References

1. Faber, F.A., Lindmaa, A., von Lilienfeld, O.A., & Armiento, R. (2016). Machine learning energies of 2 millionelpasolite (abC2D6) crystals. Phys. Rev. Lett., 117(13), https://doi.org/10.1103/ PhysRevLett.117.135502.

2. Hao, Y. Bharathi, M.S., Wang, L., Liu, Y., Chen, H., Nie, S., Wang, X., Chou, H., Tan, C., Fallahazad, B., Ramanarayan, H., Magnuson, C.W., Tutuc, E., Yakobson, B.I., McCarty, K.F., Zhang, Y.-W., Kim, P., Hone J., Colombo, L., & Ruoff, R.S. (2013). The role of surface oxygen in the growth of large single-crystal graphene on copper. Science, 342(6159), 720-723, https://doi.org/10.1126/science.1243879

3. Li, Z., Ma, X., & Xin, H. (2017). Feature engineering of machine-learning chemisorption models for catalyst design. Catalysis Today, 280:232-238, https://doi.org/10.1016/j.cattod.2016.04.013

4. Lin, J., Peng, Z., Liu, Y., Ruiz-zepeda, F., Ye, R., Samuel, E.L.G., Yacaman, M.J., Yakobson, B.I., & Tour, J.M. (2012). Laser-induced porous graphene films from commercial polymers. Nature Communications, 5, https://doi.org/10.1038%2Fncomms6714

5. Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, & K.-R., von Lilienfeld, O.A. (2013). Machine learning of molecular electronic properties in chemical compound space. New Journal of Physics, 15(9), https://doi.org/10.1088/ 1367-2630/15/9/095003

6. Nørskov, J., Abild-Pedersen, F., Studt, F., & Bligaard, T. (2011). Density functional theory in surface chemistry and catalysis. Proceedings of the National Academy of Sciences of the United States of America, 108(3), 937-943, https://doi.org/10.1073/pnas.1006652108

7. Raccuglia, P., Elbert, K.C., Adler, P.D.F., Falk, C., Wenny, M.B., Mollo, A., Zeller, M., Friedler, S.A., Schrier, J., & Norquist, A.J. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature, 533(7601), 73-76, https://doi.org/10.1038/nature17439

8. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: recent applications and prospects. npj Computational Materials, 3(54), https://doi.org/10.1038/s41524-017-0056-5

9. Rupp, M., Tkatchenko, A., Müller, K.-R., & von Lilienfeld, O.A. (2012). Fast and accurate modeling of molecular atomization energies with machine learning. Phys. Rev. Lett., 108, https://doi.org/10.1103/PhysRevLett.108.058301

10. Sendek, A.D., Yang, Q., Cubuk, E.D., Duerloo, K.A.N., Cui, Y., & Reed, E.J. (2016). Holistic computational structure screening of more than 12000 candidates for solid lithiumion conductor materials. Energy Environ. Sci., 10(1), 306-320, https://doi.org/10.1039/c6ee02697d

11. Shell, M.S. (2012). Classical semi-empirical force fields. Lecture notes for Principles of modern molecular simulation methods, University of California Santa Barbara. https://engineering.ucsb.edu/ shell/che210d/Classical_force_fields.pdf

12. Zhang, Y. & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. npj Computational Materials, 4(25), https://doi.org/10.1038/s41524-018-0081-z

---

Machine Learning and Ethnography: A Marriage Made in Heaven

Tatiana Gherman
School of Business and Economics,
Loughborough University, UK

SIMPLY stated, the essence of operations research is the creation of models to support better decision-making. Although 'modelling' is regarded as being the key term here, it is essential that we do not prioritize modelling at the expense of the 'better decision-making' element. At the end of the day, modelling that does not unlock value to improve decision-making is, in practical terms, useless.

Nowadays, in the context of the exponentially-growing data, developed models are being automated using methods of machine learning, whose applications have an enormous 'appetite' for data. It is not too bold to say that over the past years, machine learning and predictive analytics together have been revolutionizing our society by transforming the growing data into predictions that support the decision-making process (Lee, Shin, & Realff, 2018). While it might be true that applying machine learning techniques to the decision-making process can translate into a competitive advantage, a lot can go wrong along the way, especially when dealing with emergent human dynamics in the data, which can lead to inaccurate predictions.

When we speak about machine learning algorithms, we generally imagine a lot of 'crunching' of data points; but it's not just about a simple and brute computational force. From a purely cognitive computing perspective, the development of artificial models using machine learning techniques resembles the ability of human learning; in other words, it is the way to educating computers on how to perform complex tasks. The question is, can we teach algorithms to learn better? In machine learning, "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell, 1997, p. 2). Machine learning needs large datasets to learn, which implies that rather than relying on statistically relevant samples, as much data as possible is instead collected and analysed (Butterworth, 2018). In other words, "machine learning aims to build programs that develop their own analytic or descriptive approaches to a body of data, rather than employing ready-made solutions such as rule-based deduction or the regressions of more traditional statistics. They do so through repeated trials, following each of which error is identified and fed back into the system, and adjusting the approach for each subsequent trial" (Lowrie, 2017, p. 4).

In order for machine learning algorithms to automatically learn from the existing data, data is trained to reduce prediction errors and then is tested for feature extraction. But to what extent exactly can prediction errors be diminished? A quite common problem has to do with 'overfitting', which occurs when a machine learning algorithm tries too hard to hit every data point exactly, adapting itself too much to the noise in the data. It is rather obvious that understanding the data to uncover the underlying causes for the fluctuations in the data is essential; this is even more relevant if, as mentioned before, we deal with emergent human dynamics. While computational techniques are continuously being developed to address these aspects, few research efforts are actually considering the potential brought about by a different kind of approach that by excellence is able to provide deep insights into human behaviour: ethnography.

The aim of ethnography is to provide a detailed description of the phenomena under study. It involves systematic research and analysis, grounded in evidence, and it can provide insights that can lead to new hypotheses or revisions of existing theory or understanding of social life. Ethnography can offer a richer understanding of the data, of the social context from which the data comes.

Generally speaking, machine learning and ethnography are conceptualized as polar ends of a research spectrum; nonetheless, there is more common ground than is obvious at first glance – in many ways, they have a shared purpose. As noted above, machine learning algorithms can go wrong and they often do go wrong. And sometimes the reason for going wrong resides not in technical details, but in the fact that not enough effort has been dedicated to understanding the social context from which data comes. In order to develop machine learning applications that work better for society, we must be able to understand what the society looks like from inside a particular context and articulate particular stances. Together, machine learning and ethnography can provide a more comprehensive picture of data, and can generate more societal value than each approach on its own (Charles & Gherman, 2018). As of today, mixed methods research that combine machine learning and ethnographic approaches is still rather scarce; but, as the discussion about the greater good in machine learning is heating up, this type of work will grow on a greater scale. There is a scope for expanding the common ground between machine learning and ethnography.

The future of machine learning is not just about crunching more data points, but instead it is about asking deeper and more insightful questions. International Data Corporation (IDC) predicts that the digital data created worldwide will grow from 4.4 zettabytes in 2013 to 44 zettabytes by 2020 and 180 zettabytes by 2025, still there is a lot of unexplored potential. As Rattenbury and Nafus (2018) elegantly stated in a recent interview with regards to the common ground between data science/machine learning and ethnography, "[...] there's a lot of potential in collaborating to illuminate the systems that create data. Part of that potential [...] will be realized by leveraging the different epistemological assumptions behind our respective approaches. For example, there is unquestionable value in using statistical models as a lens to interpret and forecast sociocultural trends—both business value and value to growing knowledge more generally. But that value is entirely dependent on the quality of the alignment between the statistical model and the sociocultural system(s) it is built for. When there are misalignments and blind spots, the door is opened to validity issues and negative social consequences, such as those coming to light in the debates about fairness in machine learning. There are real disconnects between how data-intensive systems currently work, and what benefits societies."

References

1. Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. Computer Law & Security Review, 34, 257-268.

2. Charles, V., & Gherman, T. (2018). Big Data Analytics and Ethnography: Together for the Greater Good. In A. Emrouznejad & V. Charles (Eds.), Big Data for the Greater Good (pp. 19-34). Studies in Big Data, 42, Springer International Publishing.

3. Lee, J. H., Shin, J., & Realff, M. J. (2018). Machine learning: Overview of the recent progresses and implications for the process systems engineering field. Computers and Chemical Engineering, 114, 111-121.

4. Lowrie, I. (2017). Algorithmic rationality: Epistemology and efficiency in the data sciences. Big Data & Society, January-June, 1-13.

5. Mitchell, T. M. (1997). Machine Learning. Boston, MA: McGraw-Hill.

6. Rattenbury, T., & Nafus, D. (2018). Data Science and Ethnography: What's Our Common Ground, and Why Does It Matter? https://www.epicpeople.org/data-science-and-ethnography/

INFORMS Minority Issues Forum Spotlight

Lewis Ntaimo, Professor, Department of Industrial and Systems Engineering at Texas A&M University

THE Minority Issues Forum (MIF) has been an active and vibrant forum for INFORMS members interested in minority issues in OR/MS since 2004. The objectives of MIF are to 1) foster minority representation in OR/MS; 2) develop ties between those interested in increasing the number of minority participants in OR/MS; and 3) disseminate information about the issues that minority researchers and practitioners face. The 2018-2019 MIF board has the following membership:

Figure 5: MIF 2018-2019 board members

President - Lewis Ntaimo (Texas A&M University), *Vice President* – Ruben Proano (Rochester Institute of Technology), *Treasurer* – Jamol Pender (Cornell University), *Secretary* – Jessye Talley (Morgan State University), *Programs Co-Chair*– Karen Hicklin (North Carolina at Chapel Hill), *Programs Co-Chair* – Eduardo Perez (Texas State University), *Junior Programs Chair* – Trilce Encarnacion (Rensselaer Polytechnic Institute), *Communications Co-Chair* – Shannon Harris (Ohio State University), *Communications Co-Chair* – Laila Cure (Wichita State University), and *Immediate Past President* – Maria Mayorga (North Carolina State University). For the next two years the goal of MIF is to go beyond the progress that it has achieved so far. Specific goals include increasing the number of regular and graduate student members; improving visibility and participation of the minority and underrepresented at INFORMS; and growing the MIF financial portfolio to enable more activities. Every year MIF hosts three main events at INFORMS: 1) Student Poster competition; 2) Paper Competition; and 3) Early Career Award. MIF encourages people in the OR/MS community who are passionate about minorities issues to join the forum and participate in it's activities as judges or financial contributors to support them.

The purpose of the MIF Student Poster Competition is to highlight and promote the research of underrepresented students and their advisors. While the MIF poster session has been in place since 2004, the poster competition began in 2012. Since then, MIF has sponsored the travel of underrepresented minority graduate students who participate in the poster session and need financial support. The competition takes place during the MIF Reception at INFORMS. Final judging takes place live by a panel of distinguished volunteer "celebrity judges". The finalists are announced at the end of the reception and the first-place winner receives a monetary award. The winners of the 2017 MIF Sixth Poster Competition were Gian-Gabriel P. Garcia (University of Michigan) and Toyya A. Pujol (Georgia Tech) while Honorable Mention awards went to Donald Richardson (University of Michigan) and Lauren N. Steimle (University of Michigan).

Figure 6: 2017 Student Poster Competition Finalists

MIF started the Paper Competition in 2016 to promote and bring visibility to recent contributions of MIF members in the field of operations research, management science, or information systems. The motivation stemmed from the desire to feature work from younger scholars within the MIF population because it is often difficult to be recognized within the larger INFORMS community. Submissions must have been accepted or published within the past two years of the award year and are judged based on significance of contribution to theory or practice, novelty and technical strength of methodology, significance and clarity of results, conclusions being reasonable and supported, and organization and quality of exposition. The finalists present their work in a dedicated session at the annual conference, and the first-place winner receives a monetary award. The finalists of the 2017 MIF Second Paper Competition were Alp Akcay and Canan G. Corlu for their paper, "Simulation of inventory systems with unknown input models: a data-driven approach"; Michelle Alvarado and Lewis Ntaimo for their paper, "Chemotherapy appointment scheduling under uncertainty using mean-risk stochastic integer programming"; Esra Büyüktahtakin, and Joseph C. Hartman for their paper, "A mixed-integer programming approach to the parallel replacement problem under technological change"; and Maria Mayorga, Emmett Lodree, and Justin Wolczynscki for their paper, "The optimal assignment of spontaneous volunteers". First Place was awarded to Michelle Alvarado (University of Florida) and Lewis Ntaimo. MIF is grateful to the following who served as judges: Dr Karen Hicklin (North Carolina State University), Illya Hicks (Rice University) and Mark Lewis (Cornell University). In 2016, MIF also launched the Early Career Award. The purpose of the MIF Early Career Award is to recognize outstanding contributions to the theory or practice of OR/MS and service made by active members of MIF. The award recognizes exceptional researchers who have shown promise at the beginning of their academic or industrial career. Eligible applicants include scholars who are MIF members, pre-tenure, and within eight years of receiving their doctorate (or equivalent) degree. Individuals are either nominated or self-nominated and submit a C.V., two-page statement, and one letter of recommendation. A panel of three judges evaluate nominees based on research (potential and accomplishments to date considering intellectual merit and broader impact) and service (service to broader OR/MS community and the MIF community).

MIF is grateful and acknowledges the valuable financial support from the following universities which enabled them to carry out the above-described activities: Cornell University (ORIE), Clemson University, Georgia Tech University, University of Maryland, University of Michigan, North Carolina State University, Northwestern University, The Ohio State University (Fisher College of Business), Texas A&M University (ISEN), University of Wisconsin-Milwaukee, Copper Sponsorship, Columbia University, North Carolina A&T State University, Cornell (Johnson School of Business), and University of Pittsburgh (Katz School of Business). The announcements for the MIF activities are sent out in April and the deadline for submission is in August. The MIF contact email is mif.informs@gmail.com for those interested in participating. For more information on MIF, visit `http://connect.informs.org/minorityissuesforum/home`.

Student Chapter Spotlight: Northeastern University

The INFORMS Student Chapter at Northeastern University was established in 2015 with fewer than 50 members. Since then, our members have grown in number to 231 and have organized more than 20 events throughout 2017 and early 2018. Our student chapter aims to bring our members in closer contact with current OR and analytics issues and techniques in both academic and non-academic environments.

In Fall 2017, we organized a series of machine learning tutorials using Python taught by one of our own PhD candidates. It began with an introduction to Python and then continued bi-weekly to cover Naïve Bayes, Support Machine Vector, Decision Trees, Random Forest and Evaluation Metrics. We ended the semester with a talk by Dr. Rina Schneur, the 17th INFORMS president and former Director of Business Analytics at Verizon.

With the new board of officers elected in late 2017, the direction of our student chapter has extended to embrace more collaboration with internal and external organizations as we look to organize high- quality and proactive events. We plan to take advantage of the opportunity of being in the heart of Boston where many companies and organization headquarters are located. We have also continued to organize tutorial and seminar sessions to accommodate the needs of our members. In Spring 2018, we held two tutorials and three seminar sessions. We were delighted to have Northeastern's Data Analytics and Visualization Specialist give a tutorial on "Data Visualization with Tableau" and to have a professor from our College of Engineering provide a tutorial session on Deep Learning. For the seminar session, we invited an alumnus who is currently working in the Massachusetts Health Policy Commission to talk about designing healthcare policy and using analytics from the perspective of a practitioner. We were also proud to have Wayfair and Teradata representatives deliver seminars in which they discussed new methods and techniques in the analytics world. To encourage members to

use the data visualization techniques they learned in the tutorial sessions, we organized a "Data Visualization Hackathon." We used public data sets for Boston and allotted 6 hours for the participants to analyze the data and produce several visualizations to address the problem of study. The participants then printed their visualizations into the provided template and presented them to the judges. Winners of the hackathon received Amazon gift cards and university mugs. The program gathered 28 teams with a total of 72 participants and received great feedback to continue this event in future. We also introduced a new series of events called "Optimize your Connection" in collaboration with the Department of Mechanical and Industrial Engineering. We reserved 45 minutes to one hour for selected PhD students to meet with invited speakers so they could discuss and ask questions about the speakers' research, academic experience, and more. In the Spring 2018 series, we were honored to have Dr. Gun Udomsawat from the United States Postal Service, Dr. Jean-Francois Cordeau from University of Montreal, Dr. Karen Zheng from MIT, and Dr. Barry Nelson from Northwestern University sit down with our students.

Another new program launched this year is ProBono Analytics. We gathered a number of interested members to work on several pro bono projects. We initiated the program by meeting with non-profit organizations in a "speed-dating" environment organized by the Center of Community Service at Northeastern. We met and presented the objective of the program and received great responses from the organizations. With an intended 3-4 months project timeline, we launched 3 projects in early March with 5-6 members for each project. The program was first introduced by the larger INFORMS organization and was applied by our chapter to help our members use their problem solving and analytical skills with real-world data sets. This summer, we are looking to partner with other student chapters to create a standardized workflow for an independent ProBono Analytics program to better streamline the process so that the valuable program can help not only the non-profits but our members as well.

Our members receive a weekly newsletter in the spring and fall semesters that provides day-of event reminders as well as advertisements for upcoming events. Our newsletter also contains a section on

"Success Stories" where we announce newly published papers, awards, and other accomplishments of our members.

In the coming fall semester, we plan to organize more social events, such as Trivia Night and a game show called "JeORpardy." We are looking to continue our efforts in collaborative events with undergraduate organizations such as Northeastern University's Institute of Industrial and Systems Engineers (IISE) to promote optimization work, as well as to encourage more undergraduates to continue working in operations research and analytics. We would like to expand this collaborative effort to other student chapters, as well as learn about the events other student chapters have organized, so that we can share program content and further promote INFORMS. To subscribe to our newsletter, please email informsnu@gmail.com, and to learn more about us, follow us on Facebook and Twitter at INFORMS at Northeastern University or visit our website at `https://web.northeastern.edu/informs/`.

Student Chapter Highlights

University at Buffalo

Spring 2018 was the most eventful semester for INFORMS student chapter at University at Buffalo (UB INFORMS). We had 21 events, beating our own previous record of 15 events (Fall 2017)! These events were highly diverse, spanning apparel sale, alumni talks (from two ex-presidents of our chapter!), pro-bono analytics projects, research seminars, software/technology workshops, officer elections, and general meetings. Additionally, considering that our chapter is growing both in terms of the quality and quantity of events and in terms of engaging with the student body and the community, we have been working on improving the organizational and functional framework of our chapter. We had a meeting in the first week of May, exclusively for revising/updating our constitution/by-laws. The idea is to have such a meeting at the end of each academic year, so that constitution/by-laws are updated systematically. We celebrated the end of a successful year in the last week of the semester with a "darts-social" mix. The chapter is looking forward to another, more eventful year ahead with the newly elected officers, and also, to its first ever quarter-zip apparel that will be delivered soon!

Koç University

Since its establishment in 2000, the INFORMS Student Chapter at Koç University, the Industrial Experience Society (IES) has sought to be a family with a professional environment, while embracing the values of hard work, integrity, honesty and sincerity. The members of IES are chosen with a democratic electoral system and come from a variety of faculties, including nursing, engineering, administrative science & economics and social sciences.

The aim of IES is to assist its members and the participants of its projects on the formation of their career path. This is achieved by the presentation of several career opportunities through different activities such as LEAP, The One, Be Pro, IndEx and BOM. All of our organizations are simulations of different career paths. For instance, LEAP aims to support young entrepreneurs, whereas IndEx brings together C-Level managers of different corporations with 500 participants.

IES is member of IAESTE, INFORMS and EMT because IES values expending its relations with different communities and developing their ideas. For more information you can check out our social media platforms using the name "ieskocuni".

University of Michigan

INFORMS at the University of Michigan (UM) supports and encourages the academic, social, and professional pursuits of its student members, the Industrial and Operations Engineering Department, and wider INFORMS-related community. This past year, our chapter hosted many initiatives to serve our community (through OR, of course!). Two of our most successful events include innovative fundraising seminars which combined fundraising with state-of-the-art OR research.

The first event was Engineering Resilience, where Byron Tasseff, Dr. Sara Shashaani, and Dr. Pascal Van Hentenryck presented their work on using data-driven analytics for mitigating, responding to, and recovering from natural disasters. Fundraising efforts for this event totaled over $900 for the victims of Hurricanes Harvey, Irma, and Maria.

The second event, Movember, focused on men's health issues including prostate cancer, testicular cancer, and mental health. Participants abstained from shaving throughout November and raised a total of $1440. The event culminated in a men's and women's mustache competition and a seminar by Dr. Brian Denton on his research group's recent advances in detecting and treating prostate cancer using machine learning and optimization techniques.

Other highlights this year include our Healthcare Journal Club, International Movie Night Series, Pro-bono Initiatives, and the Data-driven NCAA Bracket Challenge.

Mississippi State University

The INFORMS Student chapter at Mississippi State University is an organization which is committed to encourage and develop interest in the fields of operations research (OR) and management science (MS). We provide a means of communication and networking among people with interests in OR/MS. We also provide an informal means of exchange about OR/MS educational programs and opportunities. We share information among students by conducting a number of activities which include Speaker Seminars, Professional Training Workshops, and Career Development Events, to list a few. In 2018, the INFORMS Student chapter at Mississippi State organized workshops on Artificial Neural Networks, MATLAB, and Supercomputing. We were lucky to have Dr. Jessica L. Heier Stamm, Assistant Professor, Department of Industrial and Manufacturing Systems Engineering, Kansas State University visit this February. She shared some of her research experiences with the members and provided guidance on ways to cope with professional challenges. We also hosted Dr. Halit Üster, Professor in the Department of Engineering Management, Information, and Systems at Southern Methodist University in April 2018. In addition to presenting his recent research, he made some valuable suggestions to make the INFORMS student chapter more successful. We are planning to organize more workshops and events both in this summer and upcoming Fall that will include the members of other on-campus student organizations. The goal is to extend the knowledge of OR and MS to other student communities and promote collaborative efforts. We would like to collaborate with INFORMS student chapters at other universities as well. We believe collaboration at a larger scale would add new dimensions to the success of the INFORMS student chapter at Mississippi State University.

Texas A&M University

The INFORMS Student Chapter at Texas A&M's primary mission is serving its community and students well. The A&M student chapter hosted a #gradTax phone-a-thon last fall where almost 100 students and faculty pledged to call/email

Texas legislators to protest the proposed tax on graduate student tuition (and hey, it worked!). We participated in both the local SPARK conference and an Engineering Open House, where our members hosted STEM activities for elementary schoolers. We are also partnering with an elementary school to support an after-school engineering club (first activity: a shortest path obstacle course!). We try to balance the needs of students pursing academic and industry careers. Our industry affairs team invited some of our graduate alumni in the healthcare, construction, retail, and chemical industries to conduct a panel on preparing for jobs in industry. We also had a Faculty Search Process panel, where various faculty and one of our postdoctoral students offered advice for navigating the academic job hunt.

This semester, we piloted student-led workshops, including crash-courses in Linear and Nonlinear Programming, Machine Learning, Integer Programming, and JMP software. But, as you know, all work and no play are no good, so we also hosted a professor-student karaoke party for a mid-semester break! Finally, we rounded out the semester with a little bit of scandal: during our end-of-semester potluck, we discovered that the winner of the "Best Dish" award used a store-bought sauce in their chicken-quinoa curry! To see what else we are working on, consider following our chapter on Facebook and Twitter.