FALL/WINTER 2014

OR/MS TOMORROW

The informs Student Magazine

Letter from the Editors

Inside This Issue

1 Letter from the Editors

2 Interview

3 Student Chapter Spotlight

4 Society Spotlight

5 Featured Research

6 OR/MS Across the Globe

7 Journal Articles of Interest

8 Conferences

9 Student Competition

Dear readers,

Welcome to the second thematic issue of OR/MS Tomorrow. We hope that you are finding the issues around a specific topic more interesting and useful. Our goal is that you will receive the latest news from the INFORMS community, as well as a picture of current research and practice in trending OR topics. Here is a brief summary of what to expect in this issue of OR/MS Tomorrow issue on OR/MS in Big Data & Analytics

We begin this article with a feature interview with Professor Michael J. Franklin, the Thomas M. Siebel Professor of Computer Science and Chair of the Computer Science Division in the EECS Department at the University of California Berkeley. Dr. Franklin addresses what makes big data an exciting area of research, some of the research challenges faced in this area, if big data is a fad, and what students should do if they are interested in studying big data and analytics. The interview will give you an understanding of this field before we move on to how OR is being used in developing countries.

In our previous issue, we focused on highlighting operations research in healthcare in South Africa. This edition continues the focus in Africa, but now focuses on big data analytics. Big data analytics has the potential to help decision makers make better decisions; help aid-workers target those most in need; and help communities fight corruption. Find out more in OR/MS Across the Globe.

Next, we continue our spotlights on a student chapter and journal. This issue's Student Chapter Spotlight is the University of Toronto Operations Research Group while the Journal/Society Spotlight is the Big Data & Society (BD&S) is a SAGE open access peer-reviewed scholarly journal that publishes interdisciplinary work principally in the social sciences, humanities and computing and their intersections with the arts and natural sciences about the implications of Big Data for societies. The Journal's key purpose is

to provide a space for connecting debates about the emerging field of Big Data practices and how they are reconfiguring academic, social, industry, business and government relations, expertise, methods, concepts and knowledge.

We end the issue with information for you to learn more about Big Data and Analytics. First, we offer access to two articles we found interesting. The first article uses LinkedIn skills database to analyze the effect of labor market factors on early returns on big data investment. The second article offers active learning methods for the use of large-margin classifiers in semantic annotation. These methods can be used to minimize the number of labelled training examples for semantic annotation tasks that are usually performed to extract business intelligence from textual data. We also summarize upcoming conference in the area of Big Data and Analytics. The conferences range from the use of big data in text mining/semantic analysis to healthcare and medicine to business analytics.

Once again, we are excited to present our second thematic issue OR/MS Tomorrow and we hope you will enjoy reading OR in Big Data and Analytics as much as we did while preparing it for you.

- The Editors (Sara, Amanda, Vaggelis, Ruixue, Andrea, and Kanshukan) ■



Interview about the Burgeoning Field of Big Data

We offer a feature interview with Professor Michael J. Franklin, the Thomas M. Siebel Professor of Computer Science and Chair of the Computer Science Division in the EECS Department at the University of California Berkeley. Dr. Franklin addresses what makes big data an exciting area of research, some of the research challenges faced in this area, if big data is a fad, and what students should do if they are interested in studying big data and analytics. The interview will give you an understanding of this field before we move on to how OR is being used in developing countries.

I recently had the opportunity to talk with Professor Franklin about the burgeoning field of big data.

Thank you so much for your time. I'm very excited to talk with you today. You've done a lot of great research in the area of big data, and I have a few questions for you about the field.

Could you tell us a little bit about what makes big data an exciting area of research?

I think the reason that big data is so exciting for two reasons. One is that it's so pervasive. If you look at all the different areas that are becoming more data-driven, it's across all human endeavors. All of the sciences are becoming data-driven: commerce, travel, health and medicine. Across the board, everything is becoming data driven. So if you work on big data, you get to work on things that can have real impact in a bunch of different areas. The other reason it's exciting is just because the technology is moving so quickly. We can do things today that just weren't even possible a few years ago.

What do you think are some of the biggest research challenges right now in big data?

There are a number of challenges that people are still working on having to do with the sheer volume of the information that's become available. So there are a number of challenges just dealing with scale and data volume.

Looking beyond that, there are a number of problems. One is, as you begin to collect data from more and more places, it becomes harder and harder to get a consistent picture of what's going on. So as you bring in more data from more places, you run into conflicts and inconsistencies, and it becomes possible to find spurious relationships that aren't really there. So dealing with data quality as you're getting data from more and more places and understanding what inferences can you safely make given certain types of data and certain collections of data. I think there's a whole interesting area that people need to explore around just understanding the quality of answers, and what you can trust and what you can't trust.

And then there are also huge issues around privacy. As data is collected on finer and finer granularities on more and more things, there's an inherent tradeoff

between what you can learn from that data and how much you're seeing about what people are doing and their personal information.

How has big data changed in the last ten years, and where do you see it going in the future?

I think the main way that it has changed is primarily its pervasiveness. Ten years ago, there were a small number of organizations that were collecting a lot of data and were using that in various ways. But now it's become democratized. Lots of organizations are collecting data, lots of companies are collecting data, and lots of scientists are collecting data. So it's just much more pervasive than it was ten years ago.

And it used to be that if you wanted to collect a lot of data and try to make sense of it, you would have to go out and build a huge data center full of expensive computers. Now you don't have to do that. Now you can go to an online, public cloud service like Amazon Web Services or Google or Microsoft and rent the computation you need for just the amount of time you need it. So really, if you just have a credit card you can get access to computing resources that ten years ago would have been available to only a few, select organizations.

Some of your work has concerned the "human side" of big data. Could you explain what the "human side" is?

In terms of the human side of big data, there are two different ways I like to look at it. One is maybe the most obvious way, thinking about an analyst or a data scientist or team of people who have a collection of data and are trying to find interesting and important insights and patterns in there. There's a big question of just how do you support the analytics process for people who are trying to find insights in data. In order to do that, you need to understand things about human perception, thought processes, and the analytical processes that people go through when they're trying to answer questions or when they're just exploring big data sets. So one side of the human equation is how do you present results and insights to people in a way that the people can understand those insights, trust them, and then know how to respond to them. Because you can produce a lot of information for people, but if they don't know how to make sense of it or they don't believe it, then it's not going to help. That's one aspect of the human side of Big Data.

The other part that's a little, maybe less obvious, is the fact that because people are increasingly connected—everyone is on the internet, everyone has access to the network through their smart phones that they carry with them all the time—you can start thinking about how do you involve a crowd in the analysis and collection of the data. You can think about having crowds of people helping you collect information. So for example, if you wanted to build an environmental sensing application, you've got millions of people running around with smart phones. Those people could help you collect data about what's going on in the environment. And then there are other platforms that have been developed where people can help solve analytical problems. Techniques of crowdsourcing and human computation can be used to aggregate people online to answer questions, to look at data, and find interesting things in the data. So the other side of the human equation is using crowd resources to collect data and make sense of data.

I'm sure a lot of students are going to be interested in your thoughts on big data and then potentially studying disciplines related to big data. If a student were interested in studying this, what recommendations would you give to him or her?

Many universities are wrestling with that exact issue right now, so it's a great question. It depends where in their studies a student is. There are a lot of new master's programs being created that are specifically aimed at data science. They have names like Master's in Data Science, Master's in Business Analytics, and Master's in Predictive Analytics. There are a bunch of programs around the country targeted at master's level people. At the undergraduate level, that hasn't quite happened yet although it's starting.

I think at the undergraduate level you need to be thinking about a combination of computer science and statistics because you really need expertise in both of those areas to be able to do effective work in big data. I would look for computer science departments that are closely related to statistics departments. For example, our department at Berkeley has a number of faculty who are jointly appointed between the computer science and statistics. Or, vice versa if you are more on the statistical or mathematical part of the spectrum, then look for programs that seem to have good relationships with the computer science departments at their schools. But there are a bunch of other areas too. There's industrial engineering. There are business schools have programs that focus on analytics. Really what you want to look for is a program that's going to give you exposure both to thinking about data in general, like you would get from a good math program or a good statistics program, and the computational

side of things. You don't want to be doing just one or the other.

Is there anything else you would like to share with our readers, and in particular, the students?

One thing that a lot of people ask me, "The big data space is so hot right now. Is this a fad? Is this going to continue? What's happening?" I think if you look at where technology is going and you look at where organizations and disciplines are going, you realize that we're just really at the very, very beginning of this. Data is becoming increasingly important, and people who understand how to work with data, at small scale and at large scale, and how to communicate the results of what is learned from working with that data have skills that are going to be in increasing demand. I don't see any slowing down of that in the near or even in the longer term.

Professor Franklin, thanks again for taking the time and doing the interview with OR/MS Tomorrow. I'm sure many of our readers will be inspired by your thoughts and will benefit from them.

Great! Thank you very much. ■

Edited by: Andrea Hupman, Industrial & Enterprise Systems Engineering, UIUC

Bio

Professor Michael J. Franklinis the Thomas M. Siebel Professor of Computer Science and Chair of the Computer Science Division in the EECS Department at the University of California Berkeley. He is also the Director of the Algorithms, Machines, and People Laboratory (AMPLab) at UC Berkeley. He earned a B.S. in Computer and Information Science from the University of Massachusetts, Amherst in 1983, a M.S.E. from the Wang Institute of Graduate Studies in 1986, and a Ph.D. in Computer Sciences from the University of Wisconsin, Madison in 1993. In addition to his faculty position, he also is currently a technical advisory board member for numerous companies, including ClearStory Data, wise.io, Databricks Inc., @Scale Inc., and Silicon Valley Data Science. He is an International Advisory Board Member for the School of Computing at the National College of Ireland and the Center for Cloud Computing and Big Data at East China Normal University. He is an editorial board member forBig Data, and was previously a member of the National Academy of Sciences Committee on Analysis of Massive Data Sets (2010-2013). Professor Franklin is an Association of Computing Machinery Fellow and has won numerous awards and grants, including the ACM SIGMOD "Test of Time" Award (2004, 2013), the IBM Faculty Award (2009), an Outstanding Advisor Award (2011), Google Faculty Research Grant (2009), IBM Faculty Partnership Award (1999, 2002, 2003), ACM Service Award (2003), Okawa Foundation Research Grant (2000), Siemens Faculty Development Award (2000), and NSF CAREER Award (1995). ■



Student Chapter Spotlight

The University of Toronto Operations Research Group

The University of Toronto Operations Research Group (UTORG) was established in Summer 2010. The rationale for such a group was to bring graduate students with research interests in Operations Research together to share ideas and knowledge under academic and social settings. Since its inception, UTORG has had proliferation of events and reached its pinnacle in Summer 2013, when it was officially recognized as INFORMS Toronto Student Chapter, the only active chapter in Canada. UTORG is also a Canadian Operations Research Society (CORS) Student Chapter in Toronto. Today, UTORG has more than 200 members including students and faculties from several southern Ontario universities.

The chapter's first academic event was hosted on September 16th, 2010 featuring Professor Michael Best, Department of Combinatorial and Optimization, University of Waterloo. After hosting over 15 such Distinguished Speaker Seminar series, it is now a bi-sessional academic event where professionals are invited from academia or industry to share knowledge with the students and faculty alike. Another activity which is among the popular events, is the biweekly Student Lunch Talk, a laid-back environment where students can bring their own brown bag lunch or just enjoy coffee and treats while discussing a research project presented by one of their colleagues. These talks provide students a platform to practice and receive feedback not only on their research but also on presentation skills. More than 40 such lunch talks have been hosted so far. UTORG also hosts a variety of technical workshops including R, Greedy algorithms, and the popular sessional LaTeX workshop by Proffessor Dionne Aleman, where new students are given basic insight into writing scientific research articles. The chapter also hosts generic workshops such as Academic Job Search, Applying to Graduate School, and the annual Teaching Workshop where faculty members or sessional instructors (including graduate students) share their teaching experience.

To unwind students from the confinements of their labs, several





Today, UTORG has more than 200 members including students and faculties from several southern Ontario universities.

periodic social events are also part of the chapter's activities. The chapter welcomes new graduate students to MIE through their Fall and Winter Social events. During fall and winter semesters, when the weather is cold in Toronto, Board Game Night is held on the last Thursday of every month. The event is open to all MIE students to play indoor board games while enjoying snacks.

A warm month is a relief in Toronto, so they take the fun to campus parks for some sports and outdoor games. UTORG also promotes healthy lifestyle to students and support worthy causes by participating in runs like Color me Rad. The next upcoming 5K run is for the Canadian Breast Cancer Foundation on October 5, 2014.

The chapter collaborates with and volunteers for several other groups at the university. It reaches out to The Operations Research Challenge (TORCH), a one-day contest where teams of high school students solve OR-related problems. It also collaborates with the MIE's undergraduate IIE group and have judged their internal competitions in the past. UTORG continuously works with MIE graduate student association



(AMIGAS) and organizes lunches with graduate students and the departmental Seminar Series speakers. The chapter takes departmental orientation as an opportunity to introduce the group to the new graduate students and also induct potential new volunteers. UTORG has several new exciting events planned out for the upcoming term including an OR mini-symposium for students to present their research in 5 minutes. You can stay up-to-date with UTORG through their website, Facebook, Twitter and Flickr. ■

Big Data & Society (BD&S)

Written by: Ville Takala, Editorial Assistant, Big Data & Society, bdseditors@sagepub.com

Big Data & Society (BD&S) is a SAGE open access peer-reviewed scholarly journal that publishes interdisciplinary work principally in the social sciences, humanities and computing, and their intersections with the arts and natural sciences about the implications of Big Data for societies.

The Journal's key purpose is to provide a space for connecting debates about the emerging field of Big Data practices and how they are reconfiguring academic, social, industry, business and government relations, expertise, methods, concepts and knowledge.

BD&S moves beyond usual notions of Big Data and treats it as an emerging field of practices that is not defined by but generative of (sometimes) novel data qualities such as high volume and granularity and complex analytics such as data linking and mining. It attends to digital content generated through online and offline practices in social, commercial, scientific, and government domains. This includes, for instance, content generated on the Internet through social media and search engines but also content generated in closed networks (commercial or government transactions) and open networks such as digital archives, open government and crowdsourced data. Rather than settling on a definition, the Journal makes this an object of interdisciplinary inquiries and debates explored through studies of a variety of topics and themes.

BD&S seeks contributions that analyse Big Data practices and/or involve empirical engagements and experiments with innovative methods that reflects on the consequences for how societies are represented (epistemologies), realised (ontologies) and governed (politics).

BD&S is a digital-only journal, and its platform accommodates a variety of multimedia to present complex images, dynamic visualisations, and video and audio content. Contents include peer reviewed research articles and colloquia, as well as sections on bookcasts, commentaries, state-of-the art demonstrations, and work by early career researchers. The Journal is open access, thus making its content widely accessible to readers. In addition, there are no article processing charges (APCs) for authors until 2017.

Founding editor, Evelyn Ruppert, discusses the aims and scope of the journal: http://youtu.be/lDYty2WGhd0

For information and news about the journal, visit bigdatasociety.net or follow @bigdatasoc on Twitter.
To submit a manuscript, visit http://mc.manuscriptcentral.com/bdas. ■

> "…interdisciplinary work principally in the social sciences, humanities and computing, and their intersections with the arts and natural sciences about the implications of Big Data for societies…"

Spotlight on OR in the developing world: Big Data Analytics in Africa

Written by: Kanshukan Rajaratnam, Department of Finance and Tax & The African Collaboration for Quantitative Finance and Risk Research, University of Cape Town, South Africa.

In the first of this series of articles on operations research in developing countries, we focused on highlighting operations research in South Africa (1). In this article, we turn our focus to big data analytics in the developing world, specifically in Africa. Big data analytics has the potential to help decision makers make better decisions; help aid-workers target those most in need; and help communities fight corruption.

Our electronic lifestyle has resulted in generation of large amount of data holding vast array of information about our lifestyles, habits, and behavior. According to SciDev.net, more data is currently created every two days than in the human history up until 2003 (2). This is "enough information to fill a stack of DVDs that would reach to the moon and back" (2). The late technology adoption, the rapid development of mobile networks and subsequent consumers' use of text messaging services and mobile internet has generated large amount of data in the developing world. This data has the potential to improve lives through targeted aid for those most in need, better business opportunities and greater availability of information; thereby improving the quality of life.

Governments, non-governmental organizations, and private industries across Africa have started using big data and analytics to improve operations. For example, epidemiologists at Harvard School of Public Health have used mobile phone data to pinpoint malaria flare-ups in order to design new measures in malaria prevention (3). In 2013, IBM signed a partnership with Waterfund LLC and the Ministry of Water and Environment in Uganda aimed to create water cost index which will aid in attracting investments for water infrastructure projects (4). IBM's provided a crucial role through the contribution of expertise in big data. In Kenya, an insurance company, UAP, has designed a micro-insurance program for small-scale farmers through the use of big data analytics to understand agriculture (5).

Despite such promising examples of the use of big data across Africa, there are many obstacles that lie in big data for development. A recent paper by UNGlobalpulse cited privacy, digital divide, access and analytical challenges as some of the concerns in this field (6). Given the often sensitive nature of data, creation and use of private data require guidelines and policy. In addition, human capital with dual expertise in big data analytics and third world problems needs to be nurtured.

Big data and its use doesn't lie in business and policy alone. A team from Africa led by South Africa bid and won a large part of the Square Kilometre Array (SKA) project, a multi-country radio astronomy project. The MeerKAT, the precursor project to the SKA, will "fill about four-and a- half million standard 4.7-gigabyte DVDs in a day" (7). This growth needs to be matched by growth in educational programs across the continent. Currently, there is a rush to develop human capacity to ensure African scientists and astronomers are empowered to extract scientific value from huge data expected to be generated through the MeerKAT and SKA project. The educational development will have positive spillover to big data analytics in business and government. However, this growth needs to be matched with growth in management expertise and understanding on the potential impact of big data analytics on business (8).

On the conference front, the theme of the 2014 Operations Research society of South Africa is Accessible Analytics. Last year, the Big Data KE Conference in Nairobi, Kenya explored how big data analytics can be exploited to business innovation. Currently, two companies with presence in Africa, Orange and Sonatel have launched a big data challenge with aims at improving people's

"…Despite such promising examples of the use of big data across Africa, there are many obstacles that lie in big data for development..."

wellness in Africa (click here for an article in IFORS for more detail). The author's personal view is that human capital development for the effective use of big data needs to increase drastically across Africa. The continent needs graduates with not only knowledge of big data analytics but also extensive knowledge of local business cultures, available resources and local problems. This requires a home-grown approach to big data analytics education. Given the US' projected need for 140,000 to 190,000 more workers with big data expertise (9), Africa cannot rely on expertise from US and the developed world alone. Universities across the region have a role to play in developing expertise within Africa. How government, policy makers, and educational institutions respond to such needs will determine the growth and success of big data analytics in the region. It must be emphasized this growth cannot be in graduate numbers alone but also growth in management expertise in the context of big data analytics. ■

References

1. https://www.informs.org/Pubs/OR-MS-Tomorrow/ORMS-Across-the-Globe
2. http://www.scidev.net/global/data/feature/obstacles-big-data-development.html#
3. http://www.technologyreview.com/featuredstory/513721/big-data-from-cheap-phones/
4. http://allafrica.com/stories/201305240016.html
5. http://www.innovationiseverywhere.com/big-data-africa-insurance-company-provides-optimized-cover-farmers-kenya/
6. http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf
7. http://www.uct.ac.za/dailynews/?id=8767
8. http://www.itnewsafrica.com/pressoffices/hitachi/bigdata.htm [1]
9. www.mckinsey.com/Features/Big_Data

Journal Articles of Interest

Big Data Analytics

In this issue of OR/MS tomorrow, we are spotlighting Big Data Analytics with two articles published during the summer.

In this issue of OR/MS tomorrow, we are spotlighting Big Data Analytics with two articles published during the summer.

A few months ago, Management Science published its first article using the term "Big Data" in its title [1]. In this article, Associate Professor Prasanna Tambe, studied how labor market factors have shaped early returns on big data investment. Dr. Tambe used data from the LinkedIn skills database and measured the employment of workers with technical skills such as Hadoop, MapReduce, and Apache Pig. A key result of this study showed that the importance of geography, corporate investment, and skill acquisition channels for explaining productivity growth differences (during the spread of new information technology innovations) is now underscored. The second article featured in this issue showcases the research methodologies in Big Data

analytics. Investigators from China and Hong Kong applied active learning methods for the use of large-margin classifiers in semantic annotation. These methods were used to minimize the number of labeled training examples for semantic annotation tasks that are usually performed to extract business intelligence from textual data. The results found a reduction of 19.74% of the number of training examples needed (compared to the best baseline method), as evaluated based on a data set from Amazon. The authors conclude that this research opens the door to the application of intelligent semantic annotation techniques to support real-world applications such as automatically analyzing consumer comments for customer relationship management.

Want to read something different in this section OR/MS Tomorrow's. Let us know at eg366@cam.ac.uk

[1] Reprinted by permission, Prasanna Tambe, Big Data Investment, Skills, and Firm Value, Management Science, volume 60, number 6, June 2014. Copyright 2014, the Institute for Operations Research and the Management Sciences, 5521 Research Park Drive, Suite 200, Catonsville, Maryland 21228 USA.

[2] Reprinted by permission, Kaiquan Xu, Stephen Shaoyi Liao, Raymond Y. K. Lau, J. Leon Zhao, When Effective Active Learning Strategies for the Use of Large-Margin Classifiers in Semantic Annotation: An Optimal Parameter Discovery Perspective, INFORMS Journal on Computing, volume 26, number 3, Summer 2014. Copyright 2014, the Institute for Operations Research and the Management Sciences, 5521 Research Park Drive, Suite 200, Catonsville, Maryland 21228 USA. ■

Edited by: Vaggelis Giannikas

Upcoming Conferences

Big Data has been under the spotlight for some years, as computing/analytical powers and storage solutions become sophisticated. Another hot topic that goes hand in hand with big data is the Internet of Things. As Michael Svilar, Managing Director of Advanced Analytics, Accenture, stated during his keynote speech at INFORMS Big Data Conference on June 24, 2014, the Internet of Things will continue to be radically changing the landscape of big data. Additionally, the intersection between big data and many other disciplines, such as healthcare and social sciences, have also been the focus of many conferences related to big data. For example, Innovation Enterprise has a series of conferences – big data and innovation summit. Below we list a few upcoming conferences related to big data and analytics.

Text mining/semantic analysis:

On December 3-4, 2014, 12th annual Text Analytics Summit West in San Francisco, CA, will focus on how companies can extract significant business insights from text analytics.

Healthcare/medicine:

On May 20-22, 2015, Big Data in Biomedicine Conference, jointly sponsored by Stanford Medicine and Oxford University, will focus on the rich repositories of biomedical information and how mining the massive amount of data can transform disease diagnosis and treatment.

Business analytics:

On April 12-14, 2015, INFORMS conference on business analytics in Huntington Beach, CA, is another well-attended INFORMS conference, which covers a variety of areas, such as healthcare, social network, and big data & analytics in general.

These conferences offer a great platform to bridge the fast growing interest in big data with different areas. Students who are interested in big data can benefit from attending these conferences by being exposed to many analytical techniques and fascinating applications. Additionally, INFORMSAnalyticsMagazine provides articles, explaining how data analytics can drive better business decisions. This bimonthly digital magazine, along withOR/MS Todayin alternating months, is another great channels for students to get a deeper understanding on big data and business practices. ■

Edited by: Ruixue Guo, Management Science and Engineering Department, Stanford University

Undergraduate Student Competition

The co-editors of INFORMS OR/MS Tomorrow invite undergraduates involved in research in OR to submit a synopsis of the work. The contributions must have been completed within the last two semesters. The best submissions will be featured in a future edition of INFORMS OR/MS Tomorrow. If interested, students should provide 1000-2000 words on the nature of the research and emphasize their contribution.

Send questions and submissions to Amanda and Sara. Submissions must be received by **March 15th, 2015** to be considered for the Spring/Summer 2015 Issue. ■