

BOLETIM

APDIO

PUBLICAÇÃO SEMESTRAL | 57 | DEZEMBRO '17

ENTREVISTA

PATRICK DE CAUSMAECKER

LUÍS SARMENTO E MÁRIO FIGUEIREDO

CIÊNCIA(S) DE DADOS:
IMPACTO E ENSINO

IO EM AÇÃO

GESTÃO DO CONSUMO
ELÉTRICO EM REDES
INTELIGENTES

ÍNDICE

03 ENTREVISTA
OR MEETS
DATA SCIENCE
Patrick De Causmaecker

06 ARTIGO DE OPINIÃO
CIÊNCIA(S) DE DADOS:
IMPACTO E ENSINO
Luis Sarmiento
Mário Figueiredo

09 TÉCNICAS DE IO
UMA BREVE INTRODUÇÃO
À DATA SCIENCE
Luís Torgo

12 IO EM AÇÃO
AGREGAÇÃO ÓTIMA PARA GESTÃO
DO CONSUMO ELÉTRICO
EM REDES INTELIGENTES
Miguel F. Anjos

14 LUGAR AOS NOVOS
CONTROLO ÓTIMO NA DIFUSÃO DE
MENSAGENS DE MARKETING VIRAL
João N.C.Gonçalves
Helena Sofia Rodrigues
M.Teresa T. Monteiro

17 O SÓCIO N.º...
...258
Maria do Carmo Guedes

18 NOTÍCIAS DA APDIO



ANA CARVALHO

Centro de Estudos de Gestão
Departamento de Engenharia e Gestão
Instituto Superior Técnico,
Universidade de Lisboa



TÂNIA RAMOS

Centro de Estudos de Gestão
Departamento de Engenharia e Gestão
Instituto Superior Técnico,
Universidade de Lisboa

EDITORIAL

Data Science é o tema do Boletim 57. Sendo este um tema de enorme relevância para a IO convidámos um leque variado de autores de modo a realçar o grande potencial desta área. No Artigo de Opinião, Luis Sarmiento e Mário Figueiredo apresentam uma introdução ao Data Science e identificam programas de ensino que devem constar nos currículos universitários, colmatando a falta de formação nesta área.

Patrick De Causmaecker, coordenador do Euro Working Group (EWG) Data Science Meets Optimization, foi entrevistado, dando-nos uma visão clara de quais as tendências nesta junção de áreas e quais as linhas de investigação futuras que têm potencial para serem exploradas.

Na rubrica Técnicas de IO, Luís Torgo apresenta o modelo CrispDM como metodologia a adotar em trabalhos de Data Science. É ainda apresentado um caso de aplicação em auditorias de deteção de fraudes.

Na rubrica IO em ação, Miguel F. Anjos apresenta uma aplicação prática de Data Science à gestão ótima do consumo elétrico em redes inteligentes. Mostrando a diversidade de aplicações do Data Science, João Gonçalves, Helena Rodrigues e M. Teresa Monteiro, ilustram a sua aplicação no controlo de mensagens de marketing viral na rubrica Lugar aos Novos. Este último trabalho foi galardoado com o Prémio EstudIO no XVIII Congresso da APDIO (IO2017).

Dando continuidade à nova rubrica "O sócio n.º...", a sorteada desta edição foi Maria do Carmo Guedes, sócia número 258, que relata a sua contribuição na IO e como é que esta

foi parte integrante de toda a sua vida profissional.

Continuando a querer envolver a comunidade IO, incentivámos o envio de notícias por parte dos sócios da APDIO. Agradecemos desde já as contribuições enviadas e apresentamos uma rubrica relativa às Notícias do segundo semestre de 2017, onde se destacam o XVIII Congresso da APDIO realizado em Valença e o reconhecimento de dois sócios da APDIO com dois títulos distintos: a Gold Medal da International Society on MCDM atribuída a José Rui Figueira e o título de EUROPT Fellow atribuído a Miguel F. Anjos. Este último semestre de 2017 foi também marcado pelo desaparecimento de Bernard Roy, fundador do EURO Working Group on MCDA, no final de outubro, o que mereceu a inclusão de uma pequena nota de homenagem nesta edição. Sendo este o nosso último boletim como editoras, resta-nos agradecer a todos os autores que contribuíram para os quatro boletins que editámos e a todos os que nos acompanharam como leitores durante estes últimos dois anos. Por último, um agradecimento especial ao Corpo Diretivo da APDIO pelo apoio e pelo voto de confiança e desejamos as maiores felicidades à próxima equipa editorial do Boletim!

Ana Carvalho Tânia Ramos

OR MEETS DATA SCIENCE

You are the coordinator of the Euro Working Group (EWG) Data Science Meets Optimization. What is the main goal of this EWG?

The EWG wants to investigate, formalize and strengthen the impact of data for decision support in operations research. Models and data are the starting point for any operations research project. The working group concentrates on optimization, which is central in operations research. Since the middle of the previous century, strong modelling techniques have been developed and are widely in use. Powerful mathematical methods as well as heuristic approaches start from a model to produce optimal, near optimal, or just 'good enough' solutions. Of course, the size of a problem instance determines the applicability of any technique. Very often, problem models do not allow for algorithms combining acceptable run time guarantees and strict optimality constraints. This is a first example of the dependence on characteristics in the data, but of course examples are ample. Performance of algorithms strongly depends on the distribution of problem instances and the model is only one part of the problem statement. If we add the dimension of time, on-line algorithms come into the picture. Changes in the data reflect the dynamism in the optimization context and this requires adaptive algorithms as well as models. The algorithm itself is a source of data when its runtime behaviour is considered. Knowledge extraction at runtime has been shown to produce significant effects in the outcomes. Another reason to make optimization and data science meet is that many data science problems are actually optimization problems. Data analytics is as essential for operations research as is optimization. The EWG wants to stimulate the exchange and application of state of the art (combinatorial) optimization techniques in data mining, machine learning as well as in any data analytic context. The long-term aim is toward a coherent theoretical framework covering the two disciplines and allowing for systematic and reliable use in applications.

You have built an impressive academic curriculum in the OR field. What are the main key challenges that you expect for the integration of Data Science with the OR methods?

Optimization models are compact descriptions, designed by optimization experts building on decennia of research. Optimization algorithms are of a mathematical nature in that they build on a deep understanding of the internal solution landscape associated with a problem. Data science builds on implicit knowledge present in sets or streams of data. This intrinsic difference between the two domains has led to a strong difference in the techniques used. Often only black box approaches are possible; black box optimization is a well-developed area. But also, automated algorithm configuration or construction, on-line algorithm selection or hyper-heuristic approaches must be considered black box.

Black box situations inevitably lead to suboptimal behaviour: opportunities to apply advanced mathematical techniques or strong diversification in search are missed, because the knowledge extracted from the data is not explicit. One possibility to improve this situation is in knowledge representation, allowing accurate representation of general knowledge in formal language. This knowledge is then used with the data to arrive at dynamic optimization models. Another possibility lies in using machine learning to learn constraints and goal functions automatically from the data. Integration with an optimization algorithm might even allow to automatically improve on model descriptions extracted from data.

When techniques are being developed, validation and testing will be necessary. One challenge is to define problems and produce good representative benchmark sets for optimization problems in which data plays an important role. Current benchmarks concentrate on the optimization algorithms; data represent challenges against which the algorithm is tested. What we need is an evaluation of the combined methodology of handling the data and the model for an optimization purpose.



PATRICK DE CAUSMAECKER

Faculty of Engineering Science, KU Leuven
patrick.decausmaecker@kuleuven.be

MODELS AND DATA ARE THE STARTING POINT FOR ANY OPERATIONS RESEARCH PROJECT

At IFORS 2017, you have chaired a session called *Data science meets optimization*. From this session, what were the main trends that you have identified?

There were actually two sessions organised on the subject. There was indeed the Data science meets optimization I chaired and there was the Data science for optimization organized by Andrew Parkes and Ender Özcan. Together these sessions aimed at setting the scene for the working group. In the 'meets' session, Emilio Carrizosa from the University of Seville made an interesting link with support vector machines by studying the cost sensitive classification problem. An interesting, and quite early, effort at creating an evaluation platform for data science and optimization was presented by Simon Wasic from Poznan University of Technology. Pieter Leyman from KU Leuven considered Optimization in Large Graphs with the ambition to cover the gap between algorithms for graphs of a few thousand vertices and those with millions as we see in social media networks. My own contribution was on the use of optimization to support the design of algorithms by innovative algorithm engineers. In the 'for' session, Andrew Parkes from the University of Nottingham discussed his work on the space of heuristics tackling a fundamental question about the structure of this space for specific problems. Morteza Davari from KU Leuven discussed schedule selection from a pre-fabricated set of feasible schedules. Joao Pedroso from the University of Porto studied orienteering on a continuous surface linking to Gaussian processes. Finally, Tu San Pham from KU Leu-

ven presented her work linking optimisation to knowledge representation.

Summarizing, Data Science and Optimization are clearly evolving toward each other, with typical data science problems such as classification and graph problems being attacked by powerful, state of the art optimization algorithms. Design and systematic support for evaluating and testing were presented through a data science based methodology and a user-friendly environment. Theoretical questions are an important subject, advances are happening and the link with knowledge representation is under investigation. Finally, two interesting examples highlighted the practical importance of applying data science to optimization. I believe these two sessions presented essential questions to be studied and discussed more deeply and extensively in the coming years. I would like to point out that there was a large stream on Data Science and Analytics. Use of optimization in analytics and use of analytics for optimization was shown useful in a number of interesting talks.

What are the applications where you see the highest potential for the integration between Data Science and Optimization?

The IFORS 2017 sessions highlighted quite a few of these. Categorization and clustering are hard problems in data science, and powerful optimization algorithms may make an important difference there. Often, algorithms already today, will work in a black box setting where the structure of the underlying data is completely unknown but supposed to be revealed by the algorithmic machinery. This may be strengthened in situations where a priori knowledge is available and can be set to use in the optimization process. Problems in graphs are another application domain for optimization algorithms. Significant literature is available in two categories; relatively small graphs where mathematical and well-known evolutionary approaches have been applied, and very large graphs where principles taken from physics are generally in use. Looking at the literature, be it significant, learns that a lot still is to be done. On small graphs, only limited methods and approaches have been studied

in detail. The larger graphs seem to suffer from discussions on the very nature of the problems. An unambiguous modelling methodology, mapping real world questions to graph problems is still lacking. Finally, little work has been done on the intermediate range, where optimization algorithms should have potential. The availability and accessibility of data allows configuring algorithms to optimally meet their daily task. This can be considered off line, where historic data are processed before going into production, and on line where the algorithms learn and improve while being used for problems popping up. The difference is in the evolution of the problem instance distribution and its effects are presently under study. Effectively, this will render the algorithms independent of the real-world problem environment and make them parameter free. In the extreme, an algorithm can be designed to dynamically learn about a problem while executing, simultaneously trying to optimize its immediate performance and to learn for future optimization tasks. Or it can be constructed from a data set as

OPTIMIZATION ALGORITHMS ARE OF A MATHEMATICAL NATURE IN THAT THEY BUILD ON A DEEP UNDERSTANDING OF THE INTERNAL SOLUTION LANDSCAPE ASSOCIATED WITH A PROBLEM. DATA SCIENCE BUILDS ON IMPLICIT KNOWLEDGE PRESENT IN SETS OR STREAMS OF DATA

CATEGORIZATION AND CLUSTERING ARE HARD PROBLEMS IN DATA SCIENCE, AND POWERFUL OPTIMIZATION ALGORITHMS MAY MAKE AN IMPORTANT DIFFERENCE THERE

is presently being studied in domains of automated construction and the different lines of hyper heuristics. The automated setting works without human assistance. Until today however, the most powerful algorithms have a human designer as the first initiator and automated algorithm construction often works best with the algorithm components invented by humans in the first place. Data science can offer assistance to such inventors, provided technical interference with the creative process and the associated noise can be kept to a minimum. Strong development environments supporting the use of data sets and making use of powerful computer infrastructure is an important line of thinking.

Another important application domain will be the case where explicit information on constraints as well as goals is only implicitly available in the data. Think e.g. about systems offering services or recommending goods to customers. Apart from preferences and vague correspondences as presently being used, hard constraints and consecutiveness conditions may apply without the system realizing. Considering these is bound to improve the performance of the system. Think about timetabling applications where certain work schedules are

considered infeasible by specific workers because of specific personal conditions, but do not explicitly state for whatever reason. Systems able to infer such hard constraints may be superior at producing optimal solutions because the optimization algorithms are able to exploit the new knowledge. We are thus making links with a large number of research domains within data science, and even more general, artificial intelligence.

Integration between data science and optimization sets a new paradigm for the two domains. Data science will profit from modern optimization techniques only if it assimilates its very strong modelling techniques. Optimization will profit from data science only if it adopts to data analytics, both implicitly through empirical, sampling based, black box techniques and explicitly through formal extraction and representation of knowledge. In both cases algorithms as well as modelling will be effected. We may need fundamentally different representations and descriptions of problems in the two domains in a new theoretical framework supported by tools in a practical setting.

When we start combining different fields, the necessity for different technical skills arises. What are the recommended backgrounds, for the human resources, in order to achieve a high-quality team working on this integration?

I require my students to take classes in advanced optimization techniques, mathematical as well as heuristic, leading to a deep understanding of what 'optimization' means, which options exist for modelling and how intelligent efficient methods work. They do need to understand the theoretical concepts. On the data side, they take advanced machine learning and data mining courses. Some take knowledge representation, some go to big data or advanced artificial intelligence and probabilistic modelling of uncertainty. This creates a solid background and a starting point for understanding the other domain. At present, students still mainly come with an operations research background from their masters, be it rooted in computer science or mathematics. We do need input from an artificial intelligence

or data analytics background and therefore we presently have a close collaboration with the group on declarative languages and artificial intelligence. There we look for students with backgrounds in data science and let them take classes in optimization. This creates the opportunity to exchange ideas at postdoc and higher levels. The domains are presently still too far apart to integrate the required knowledge and abilities in one profile. Progress must come from collaboration and exchange.

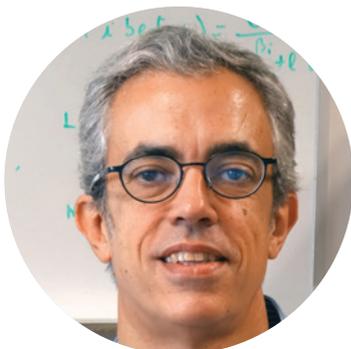
THE DOMAINS ARE PRESENTLY STILL TOO FAR APART TO INTEGRATE THE REQUIRED KNOWLEDGE AND ABILITIES IN ONE PROFILE. PROGRESS MUST COME FROM COLLABORATION AND EXCHANGE

CIÊNCIA(S) DE DADOS: IMPACTO E ENSINO



LUIS SARMENTO

Tonic App, Porto.
luis.sarmiento@gmail.com



MÁRIO FIGUEIREDO

Instituto de Telecomunicações e Instituto Superior Técnico, Universidade de Lisboa.
mario.figueiredo@lx.it.pt

1. INTRODUÇÃO

O termo *Data Science* (Ciência de Dados - CdD) é habitualmente usado para designar o estudo de problemas associados à extracção, a partir de dados (possivelmente, em grandes volumes – *big data*), de informação útil em algum sentido (comercial, económico, político, social, científico, médico) e também à aquisição, armazenamento, gestão e organização destes dados.

O interesse das organizações (privadas e públicas) na CdD é hoje enorme. As empresas conhecem bem o valor económico/comercial da informação e da sua utilização como elemento diferenciador entre competidores. Um estudo recente [1] mostra que as empresas que estão no terço superior das respectivas áreas na intensidade de utilização de decisões com base em CdD são em média 6% mais lucrativas do que a restantes. O sector financeiro usa intensamente a CdD no suporte a decisões estratégicas, entre outras aplicações. A CdD é central na segurança moderna (informática e não só), pois só analisando os enormes volumes de dados que fluem nas redes de comunicação se podem prever e identificar ameaças.

Este crescente interesse na CdD reflecte-se num aumento das oportunidades de trabalho na área. Grandes empresas globais (Amazon, Facebook, Google, IBM, LinkedIn, Microsoft, etc) e uma miríade de pequenas e médias empresas procuram avidamente *data scientists*. Esta tendência verifica-se também em Portugal, quer em grandes empresas, quer nas *startups* tecnológicas.

Podem identificar-se dois grandes catalisadores para o crescimento exponencial do impacto real e da visibilidade pública da CdD:

1) A expansão explosiva da *Internet* permitiu a partilha e o acesso a dados e informação numa escala inimaginável. Mais de 50% da humanidade tem hoje acesso à *Internet*, quer como consumidor de informação e comunicação, quer como produtor de conteúdos (texto, imagens, vídeos). Esta expansão suporta-se em (e estimula) grandes avanços tecnológicos no armazenamento, comunicação e processamento de dados (nomeadamente, a “cloud”) e nas telecomunicações, incluindo os ubíquos *smart phones*.

2) Os desenvolvimentos metodológicos, computacionais e teóricos, nas áreas da aprendi-

zagem automática (*machine learning*), inteligência artificial e visualização, alicerçados em optimização, estatística e outros ramos da matemática, permite lidar com enormes quantidades de dados, extraindo informação de forma cada vez mais eficaz.

A CdD tem também presença e influência crescentes em muitas áreas da ciência moderna, tais como biologia, astronomia, física, economia, sociologia, linguística e outras. Muitas das questões abordadas nestas áreas envolvem hoje a aquisição e análise de enormes volumes de dados, tornando a CdD uma ferramenta essencial da ciência.

É vital para o ensino superior português investir na formação em CdD, como está a suceder em muitos outros países. Sendo uma área vasta, multidisciplinar e em rápida evolução, o seu ensino coloca questões e desafios complexos. Debruçar-nos-emos sobre estas questões, após uma breve revisão histórica que dará pistas para o posicionamento da academia portuguesa em relação a esta área tão dinâmica.

2. PERSPECTIVA HISTÓRICA

Até ao início do sec. XXI, grande parte do esforço tecnológico em torno da *Internet* e da *Web* focava-se na disponibilização, em grande escala, de dados em diversos formatos. Nessa altura, o simples acesso a conteúdos era um valor em si; a indústria crescia em torno da criação da infraestrutura para produzir, codificar, armazenar e transmitir esses conteúdos. Surgiram operadores de rede, portais *Web* e empresas especializadas na codificação-des-codificação de media. A academia portuguesa acompanhou este desenvolvimento através da expansão e criação de novos cursos e/ou disciplinas de engenharia informática, redes, telecomunicações, multimédia e outras.

Na primeira década deste século, consolidada a infraestrutura tecnológica, o fornecimento de conteúdos perdeu parte do valor económico: quando o conteúdo abunda, o valor transfere-se da produção para a pesquisa e seleção. Neste contexto, ganharam preponderância duas tecnologias que serão sementes da CdD: pesquisa *Web* e sistemas de recomendação. Paralelamente, o modelo de negócio na *Web* foi-se baseando cada vez mais na colocação de anúncios em paralelo com os conteúdos. Com o aparecimento das redes sociais, em

O INTERESSE DAS ORGANIZAÇÕES (PRIVADAS E PÚBLICAS) NA CdD É HOJE ENORME. AS EMPRESAS CONHECEM BEM O VALOR ECONÓMICO/COMERCIAL DA INFORMAÇÃO E DA SUA UTILIZAÇÃO COMO ELEMENTO DIFERENCIADOR ENTRE COMPETIDORES

que consumimos e, simultaneamente, produzimos conteúdos, criaram-se as condições ideais para as empresas do triunvirato “pesquisa-recomendação-publicidade” (nomeadamente Google, Facebook e, de certo modo, a Amazon) florescerem, para desgraça das que tinham apostado na produção de conteúdos (por exemplo, empresas com origem no media tradicionais, como jornais e televisão).

O funcionamento das empresas de pesquisa-recomendação-publicidade baseia-se em, através de múltiplas formas de interacção, extrair o máximo de conhecimento acerca de cada um de nós para o explorar comercialmente. É neste contexto de guerra tecnológica pela nossa atenção que se expandem as tecnologias de “analytics”, “data mining” e “big data” para o processamento e análise de dados de utilizador, suportadas num rápido avanço da aprendizagem automática (*machine learning*) e estimulando esse avanço.

Num ambiente tecnológico diverso e ainda imaturo, a indústria absorveu pessoas com competências híbridas, da engenharia pura à análise de dados. Em geral, a academia – não só portuguesa – teve dificuldade em acompanhar as necessidades de formação nesta área, pois a mudança de paradigma no seio da indústria foi rápida e orgânica. Só mais recentemente os cursos de informática começaram a mover o foco dos seus curricula das tecnologias de informação (vaga anterior) para tópi-

cos de processamento e análise de dados.

Em plena efervescência competitiva por quota de mercado (e recursos humanos), o acesso às ferramentas de processamento e análise de grandes volumes de dados liberalizou-se radicalmente. Se, em 2005, estas eram usadas quase exclusivamente por especialistas num punhado de empresas, a partir de 2010, a próprias empresas que as desenvolveram disponibilizam-nas, bem como o acesso às respectivas infraestruturas computacionais (*cloud computing*).

Nesta altura, o sucesso económico desta indústria global contrastava fortemente com o estrangulamento das opções de carreira na academia, motivando um fluxo de doutorados para a indústria. Esta injeção de capital humano qualificado fez surgir em muitas empresas, e no seio de equipas de produto, pequenas sub-equipas de exploração de dados, as quais acabariam por ficar conhecidas como equipas de *data science*. Para lá de *analytics* ou *business intelligence*, áreas já estabelecidas, o que era pedido a estes *data scientists* era que “construíssem o futuro”: usando análise de dados, deviam conseguir criar protótipos de novas funcionalidades ou mesmo novos produtos.

Com a consolidação do papel do *data scientist* na indústria, tornou-se óbvio que o seu perfil de formação diferia dos existentes. Por um lado, para ser efectivo e autónomo, necessita de conhecimentos de engenharia. Por outro, como cada vez mais o produto se suporta em (ou se confunde com) dados, tem de ter formação sólida em processamento e análise de dados. E, em pleno ambiente empresarial, um *data scientist* deve ter também elevada sensibilidade para o produto/negócio, para orientar o seu trabalho nas direcções de maior potencial. F inalmente, fazendo jus à designação “scientist”, deve ter uma abordagem científica, criando processos para avaliar e validar os resultados da sua actividade. Em conclusão, este perfil integra traços de criativo de produto e de engenheiro, mas integra também componentes metodológicas típicas de trabalho científico.

3. ENSINO E FORMAÇÃO EM CdD

Pode pensar-se a formação em CdD a três níveis – mestrado integrado, pós-graduação profissional, doutoramento – com âmbitos e públicos-alvo diferentes. A característica comum,

e diferenciadora relativamente a ofertas existentes, é a necessidade de cobrir não só áreas formais e técnicas como também as ligadas ao desenvolvimento de produto, metodologias de avaliação e estratégias de *data-driven problem-solving*. Daremos especial ênfase ao mestrado, por ser o que pode dotar os alunos da formação base de mais largo espectro.

3.1 Mestrado

Há uma clara necessidade de um mestrado próprio e original em CdD. Embora existam já muitos módulos de outros cursos que teriam claramente lugar num mestrado em CdD, julgamos haver necessidade de criar alguns de raiz. Dado que a CdD teve, em parte, origem na indústria, é importante criar fios condutores entre as bases teóricas/técnicas e a experiência prática de resolver problemas de e com dados. Julgamos que um curso de mestrado em CdD deverá ser organizado em 5 grandes áreas:

a) Aspectos formais. Diferentes áreas de aplicação lidam com diferentes tipos de dados, mas um *data scientist* deve compreendê-los a um nível de abstracção superior: séries temporais, sequências de símbolos (por exemplo, texto), grafos, listas de tuplos, tensores, etc. Deve dominar a matemática, as estruturas de dados e os algoritmos para a sua manipulação. Isto requer disciplinas fundamentais de matemática e ciência de computadores, possivelmente todas elas já existentes em cursos actuais, mas que têm de ser integradas de forma coerente.

b) Manipulação de dados. Na prática, um *data scientist* deve ser capaz de testar ideias rapidamente, com autonomia, o que será impos-

EMPRESAS QUE ESTÃO NO TERÇO SUPERIOR DAS RESPECTIVAS ÁREAS NA INTENSIDADE DE UTILIZAÇÃO DE DECISÕES COM BASE EM CdD SÃO EM MÉDIA 6% MAIS LUCRATIVAS DO QUE AS RESTANTES

É CRUCIAL QUE A SOCIEDADE, COMO UM TODO, ESTEJA ATENTA, NÃO SÓ AO POTENCIAL DA CdD, MAS TAMBÉM AOS RISCOS QUE O SEU USO PODE IMPLICAR. PARA LÁ DAS QUESTÕES DE PRIVACIDADE, PARA AS QUAIS COMEÇA A HAVER CONSCIENCIALIZAÇÃO, AS RECENTES NOTÍCIAS ACERCA DO MODO COMO PODEM TER SIDO INFLUENCIADAS AS ELEIÇÕES NOS EUA ATRAVÉS DO USO DE TÉCNICAS CdD, DEVEM FAZER-NOS REFLECTIR

sível sem dominar métodos de manipulação (eficiente) de dados. Isto envolve disciplinas de programação, mas também outras que abordem bases de dados, tecnologias *big data*, computação de alto desempenho e distribuída.

c) Modelização, processamento e análise.

Estes tópicos são, tipicamente, cobertos em disciplinas de estatística, processamento de sinais, aprendizagem automática, investigação operacional. Num contexto de CdD, deve dar-se ao aluno um mapa actual destas áreas tão vastas, com pontos de referência e formação de base para que possam orientar a sua posterior auto-aprendizagem. É desejável que a exposição seja ilustrada com diferentes aplicações (por exemplo, finanças, sistemas de recomendação, medicina), enfatizando o carácter “todo-o-terreno” destas técnicas e tornando

claras as semelhanças formais entre os vários domínios, quando vistos a um nível abstracto.

d) Resolução de problemas. Um dos aspetos chave do sucesso de um *data-scientist* é a capacidade de pensar os problemas em função do seu contexto. Parece-nos essencial um conjunto de módulos que abordem: modelos de negócios comuns, seus objetivos e métricas; modelos de organização interna na indústria, com ênfase em equipas de produto e I&D; estratégias de prototipagem e de organização do trabalho para avaliação rápida e progresso consistente. Sendo estes módulos práticos, poderão ser assegurados em colaboração com parceiros externos de diversos domínios (empresariais e não só).

e) Metodologias de avaliação e experimentação. Uma das principais responsabilidades de um *data scientist* é o uso de metodologias científicas, nomeadamente na avaliação sistemática de soluções propostas. Assim, uma componente importante e diferenciadora de um mestrado em CdD será cobrir questões metodológicas e técnicas de avaliação diversas (por exemplo, *randomized trials*).

Integrar todas estas componentes num curso único é um desafio a encarar, resistindo à tentação de ver um mestrado em CdD como uma recombinação de tópicos já existentes. A CdD está, claro, assente em muitos tópicos clássicos mas é também uma forma moderna e original de pensar e abordar problemas práticos.

3.2 Pós-graduações e Doutoramentos

Dado a enorme procura de *data scientists*, muitos profissionais já no mercado de trabalho procuram formação neste domínio. No sentido de servir esta procura, e aproveitando a falta de resposta célere das universidades, têm surgido “academias” de CdD que, frequentemente, se limitam a exemplificar o funcionamento de ferramentas computacionais. Há, claramente, espaço para as universidades terem um papel activo na formação pós-graduada em CdD. As áreas de formação acima enumeradas podem, individualmente ou combinadas, ser adaptadas a cursos de formação para profissionais no activo.

Relativamente a doutoramentos, a situação parece-nos menos clara. Questões matemáticas,

de modelação ou algorítmicas são naturalmente enquadradas em respectivas áreas clássicas em que se inserem. As questões científicas mais específicas da CdD estão relacionadas com metodologias de abordagem de problemas, de avaliação e experimentação. Pode argumentar-se que a CdD está ainda muito embrionária para se perceber claramente que questões faz sentido abordar a nível doutoral. Essas questões emergirão provavelmente da prática, pelo que caberá à academia manter-se atenta aos desenvolvimentos da indústria, e à indústria saber interagir de forma aberta com a academia.

4. COMENTÁRIO FINAIS

É consensual que a CdD tem um enorme impacto económico directo através do seu uso nas empresas. Com a crescente automatização de muitas áreas da vida pessoal e social, a importância da CdD e o seu peso na economia continuarão a crescer. Para além do óbvio impacto económico, pode-se dizer que a CdD representa uma democratização e uma vitória da ciência empírica. Após séculos de refinamento em círculos restritos, é interessante ver as componentes chave do método científico (recolha de dados, experimentação, avaliação e validação de hipóteses) tornaram-se na base de grande parte da economia moderna.

Não podemos terminar sem referir como é crucial que a sociedade, como um todo, esteja atenta, não só ao potencial da CdD, mas também aos riscos que o seu uso pode implicar. Para lá das questões de privacidade, para as quais começa a haver consciencialização, as recentes notícias acerca do modo como podem ter sido influenciadas as eleições nos EUA através do uso de técnicas CdD, devem fazer-nos reflectir. Há que reconhecer que o estudo das implicações sociais e políticas da CdD não é da responsabilidade exclusiva dos especialistas da área, mas sim uma obrigação da sociedade como um todo. É assim importante que várias áreas, nomeadamente das ciências sociais e humanas (sociologia, ciência política, antropologia, psicologia) se debruçem sobre os impactos da CdD. Mais uma vez, caberá à academia ajudar a criar condições que facilitem esse estudo e estimular uma reflexão informada (não alarmista) na opinião pública.

REFERÊNCIAS

[1] McAfee & Brynjolfsson, “Big data: the management revolution”, *Harvard Business Rev.*, 2012.

UMA BREVE INTRODUÇÃO À DATA SCIENCE

Poucas pessoas ignorarão que vivemos num mundo em que uma grande percentagem das atividades humanas são de alguma forma registadas em formato digital. Este facto irrefutável levou a uma profusão de bases de dados que “escondem” informação potencialmente útil para as organizações envolvidas nessas atividades. Ser capaz de aceder, explorar, analisar e extrair informação deste volume gigantesco de dados tornou-se uma vantagem competitiva que estas organizações deixaram de poder ignorar. Diversos avanços tecnológicos ajudam a explicar este crescimento exponencial dos dados disponíveis. Tecnologias como sensores de todo o tipo a preços cada vez mais baixos, o poder computacional cada vez maior, a omnipresença da Internet, ou a mais recente “Internet of Things (IoT)” com inúmeros dispositivos comuns (p.ex. eletrodomésticos, automóveis, etc.) a estarem interligados e a gerarem dados, são algumas das causas deste crescimento exponencial. Independentemente das causas, o facto é que vivemos rodeados de dados que “escondem” informação potencialmente útil sobre as atividades que os geram. Ser capaz de descobrir esta informação tornou-se uma prioridade para a maioria das organizações, sob o risco de se tornarem rapidamente obsoletas. Não é, pois, surpreendente que já em Outubro de 2012 um artigo na revista “Harvard Business Review” tenha etiquetado a profissão de “Data Scientist” como “the sexiest job of the 21st century”. Vivemos pois no tempo da *Data Science*, e isto deveria ser refletido não só na indústria, mas também na academia que deveria estar a investir fortemente na formação deste tipo de profissionais que têm e terão forte procura no mercado de trabalho.

O QUE É A DATA SCIENCE?

Sobre o que estamos exatamente a falar quando usamos o termo *Data Science*? Na realidade este termo não é muito mais do que a *buzzword* que a indústria usa atualmente para qualificar atividades que não são muito diferentes da muito antiga análise estatística de dados. Antes deste termo, outros foram usados, como *business analytics*, *analytics*, *data mining*, ou *knowledge discovery from*

databases, e certamente outros se seguirão. De qualquer forma o que estamos a falar é mesmo da análise de dados. Mas será que estamos só mesmo a re-inventar a roda e a tentar vender com um novo nome as metodologias de análise estatística de dados? Se existe algum fundo de verdade nesta questão (e em muitos casos é mesmo isso que se passa), a realidade é que os dados mudaram e com essa mudança foi necessário também mudar muitas abordagens e técnicas para lidar com eles, o que de alguma forma justifica estes novos termos. E o que é que mudou assim tanto? Vários fatores, a começar pelo tamanho. Hoje em dia toda a gente gosta de falar de *Big Data*, e embora muitas vezes seja mesmo uma questão de modas, noutras estamos de facto a lidar com um volume de dados, com uma velocidade de geração de novos dados e com uma variedade do tipo de dados, que de facto não permitem a utilização direta das nossas ferramentas tradicionais de análise de dados. Longe vão os tempos em que os dados sobre a atividade da nossa organização se resumiam a uma pequena tabela numa qualquer folha de cálculo. Hoje temos sensores em tempo real a produzir informação útil, temos vídeos, temos imagens, temos documentos de texto, temos som, enfim temos toda uma profusão de formatos, fontes e tipos de dados, que complicam sobremaneira a tarefa de extrair informação útil dos mesmos.

O PROCESSO DE DATA SCIENCE

A complexidade e dimensionalidade dos dados que são analisados hoje em dia levou a um aumento da complexidade do processo de extração de informação dos mesmos. Neste contexto, o *workflow* de um qualquer projecto de *Data Science* tornou-se mais complexo, envolvendo diferentes etapas e frequentemente equipas grandes.

O modelo *CrispDM* (ver Figura 1) é um dos *standards* mais frequentemente usado para descrever os principais passos de um projecto de *Data Science*. Este modelo consiste num processo cíclico formado por seis passos principais, cujos objetivos descrevemos em seguida brevemente. O primeiro passo, *Business Understanding*, é crítico no sentido que requiere competências ao nível de *soft*



LUÍS TORGO

Departamento de Ciência de Computadores
Faculdade de Ciências, Universidade do Porto
ltorgo@dcc.fc.up.pt

skills que frequentemente só conseguem ser adquiridas com muita prática. É neste passo que a equipa de análise irá ter o primeiro contacto com o problema e com o “cliente”. É também aqui que nos começamos por debater com os frequentes problemas de diferentes contextos e linguagens. De facto, não é raro que haja uma grande diferença entre a linguagem, o conhecimento e as expectativas dos clientes, e os da equipa de analistas. Ser capaz de estabelecer pontes que tornem esta comunicação inicial eficiente é uma “arte” que não é fácil dominar. Os principais objetivos deste passo crítico são: (i) entender o negócio (qual o contexto, quais os objetivos, quais os critérios de sucesso); (ii) aferir o estado atual (que recursos existem, quais os riscos, quais os custos e benefícios, qual a terminologia usada); (iii) decidir quais os objetivos e métricas de sucesso para o projecto de *Data Science*; (iv) produzir um primeiro plano para o projecto.

O segundo passo do processo *CrispDM* é o *Data Understanding*. Este passo envolve: (i) reunir os dados disponíveis; (ii) descrever os mesmos; (iii) explorar os dados (por exemplo usando ferramentas de visualização ou

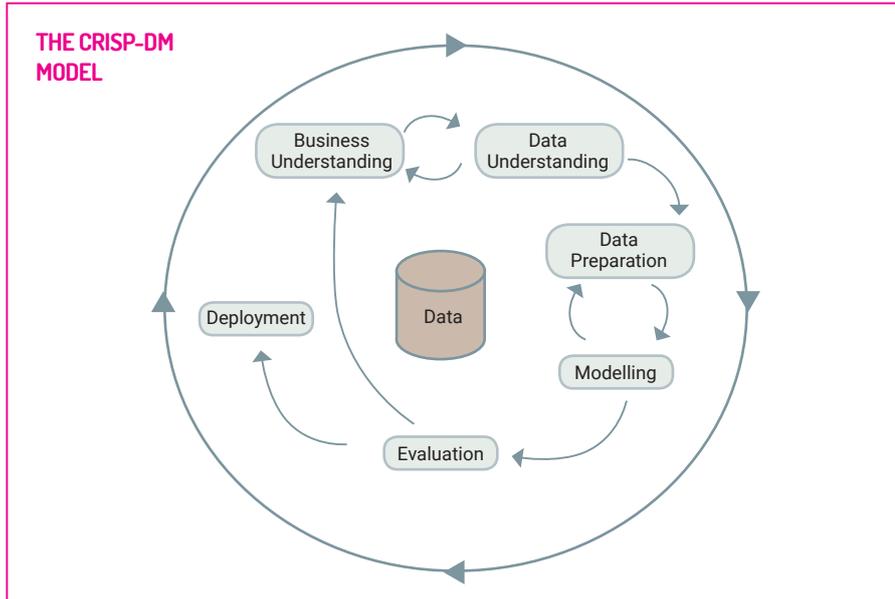


Fig. 1 - Crisp-DM model.

outras); (iv) e verificar questões relacionadas com a qualidade dos dados, uma vez que na maioria das aplicações reais de *Data Science* é frequente a existência de problemas e/ou erros nos dados.

O resultado deste segundo passo vai alimentar o passo seguinte, *Data Preparation*. Aqui o objetivo principal é a criação do conjunto de dados que vai ser usado na construção de modelos. Isto envolve: (i) a seleção do sub-conjunto dos dados a usar; (ii) a limpeza desses dados; (iii) a eventual criação de novas variáveis que possam ajudar a análise; (iv) e a eventual junção de dados de diferentes fontes.

Segue-se o passo de Modelação, frequentemente considerado um dos mais importantes, importando, todavia, realçar que, normalmente, a maior parte do tempo de um projeto de *Data Science* é na realidade ocupado pelos passos anteriores de entendimento e preparação dos dados. O passo de Modelação envolve: (i) selecionar as técnicas a aplicar aos dados; (ii) planear as experiências a levar a cabo; (iii) obter os modelos; (iv) avaliá-los e eventualmente tentar novas parametrizações dos mesmos em função dessa avaliação.

O passo seguinte consiste na Avaliação. É

nesta fase que vamos alinhar os resultados da avaliação dos modelos com os objetivos de negócio. Isto pode levar à revisão de todo o processo se os resultados não forem satisfatórios ou os esperados. Finalmente esta fase de Avaliação irá levar à decisão quanto aos passos seguintes, bem como a uma decisão em termos da passagem para ambiente de produção, ou não, dos modelos desenvolvidos.

No caso de decisão positiva, segue-se o último passo de passar o processo para ambiente de produção, ou seja, integrar os modelos desenvolvidos no processo atual da organização cliente.

Com maiores ou menores variantes, o processo *CrispDM* tem a vantagem de chamar a atenção para os principais passos e dificuldades da implementação de um processo de *Data Science* numa qualquer organização. Obviamente muitos destes passos são dependentes do contexto particular das organizações, mas a maioria das questões descritas neste processo irão certamente aparecer.

UMA ILUSTRAÇÃO DA APLICAÇÃO DE *DATA SCIENCE*

Pela natureza da motivação para a *Data Science*, não surpreende que as áreas de

aplicação desta ciência sejam bastante abrangentes. Ainda assim, há naturalmente áreas em que pela abundância de dados, seja fácil encontrar aplicações deste tipo de técnicas. Uma destas áreas é a deteção e prevenção da fraude, sendo que este tipo de atividades podem ter lugar em inúmeras aplicações concretas, desde as telecomunicações, às transações com cartões de crédito, etc.. Em seguida apresentamos uma pequena ilustração da aplicação de *Data Science* a esta classe de problemas, em particular em situações em que os recursos de auditoria disponíveis para a tarefa de deteção de fraude, são limitados.

As fraudes são, por definição, desvios à normalidade. Neste contexto, não surpreende que sejam normalmente abordadas utilizando as técnicas de deteção de *outliers*, uma vez que uma definição informal de *outlier* é exatamente uma observação que se desvia tanto do que é comum observar ao ponto de gerar fortes suspeitas de que foi gerada por um mecanismo diferente. A grande maioria dos métodos existentes para deteção de

EM OUTUBRO DE 2012, UM ARTIGO NA REVISTA "HARVARD BUSINESS REVIEW" ETIQUETOU A PROFISSÃO DE "DATA SCIENTIST" COMO "THE SEXIEST JOB OF THE 21ST CENTURY". VIVEMOS, POIS, NO TEMPO DA DATA SCIENCE, E ISTO DEVERIA SER REFLETIDO NÃO SÓ NA INDÚSTRIA, MAS TAMBÉM NA ACADEMIA QUE DEVERIA ESTAR A INVESTIR FORTEMENTE NA FORMAÇÃO DESTE TIPO DE PROFISSIONAIS QUE TÊM E TERÃO FORTE PROCURA NO MERCADO DE TRABALHO

outliers produzem, para cada caso, uma resposta binária: sim ou não é um *outlier*. Este tipo de resposta não é o mais adequado para contextos de auditoria com recursos limitados. De facto, nestas aplicações, e dado um conjunto (frequentemente vasto) de candidatos a auditoria, normalmente não podemos auditar todos os elementos deste conjunto pelo que qualquer método que ignore esta limitação de recursos de auditoria corre o risco de indicar como *outliers* mais casos dos que conseguimos auditar. Neste tipo de aplicações, mais do que ser capaz de indicar todos os *outliers*, nós procuramos métodos que maximizem a utilização dos nossos recursos limitados de auditoria. Em muitas aplicações reais de deteção de fraude é possível associar um custo à atividade de inspeção, bem como calcular o benefício da mesma, no caso de ter sido confirmada a fraude. Nestas aplicações, o método ideal é aquele que maximiza a utilidade económica das atividades de auditoria [1]. Quer isto dizer que poderemos mesmo preferir inspecionar um caso A cuja probabilidade estimada de ser uma fraude é menor do que um caso B, desde que a utilidade esperada, se confirmada a fraude, for maior em A do que em B. Formalmente, podemos definir a utilidade esperada de auditar um caso i como:

$$E[U_i] = P_i \times (B_i - C_i) + (1 - P_i) \times (-C_i)$$

em que P_i é a probabilidade estimada do caso i ser uma fraude/*outlier*, B_i é o benefício estimado se i for confirmado como fraude, e C_i é o custo estimado da inspeção de i .

Para podermos aplicar esta metodologia precisamos não só de um método de deteção de *outliers* que nos forneça o valor de P_i , mas também precisamos de obter previsões para os valores B_i e C_i . Estes valores podem ser obtidos assumindo que temos um registo histórico de atividades de auditoria passadas, para as quais registamos as características dos casos auditados, bem como o respetivo benefício e custo dessa auditoria. Com base nestes registos históricos, utilizando metodologias existentes de aprendizagem supervisionada, podemos obter modelos que, dadas as características de um novo caso, candidato a ser inspecionado, nos

possam fornecer as estimativas dos valores B_i e C_i .

Usando esta metodologia de deteção de fraude baseada em utilidade, dado um conjunto de casos candidatos a ser auditados, podemos obter uma estimativa da utilidade de auditar cada um deles e assim produzir um *ranking* de auditoria que pode ser usado para maximizar o retorno das atividades de auditoria para os recursos (limitados) disponíveis.

DATA SCIENCE: QUE FUTURO?

Por todas as razões já descritas anteriormente, o futuro da *Data Science* será certamente risinho. Existem todavia bastantes desafios que importa considerar. Do lado da academia existe uma clara necessidade de alertar os agentes para a urgência em rapidamente formar profissionais nesta área em função da pressão do mercado de trabalho. A oferta atual ainda é muito curta, embora novas iniciativas estejam na forja. No entanto, este não é um processo fácil. Os recursos humanos com competências na área são escassos e muitos estão a ser aliciados pela indústria com um poder económico com o qual a academia deste país não pode claramente competir. Acresce que não podemos olhar para isto numa escala nacional e quando falamos na indústria estamos a falar de *players* de todo o mundo que vêm em Portugal recursos humanos com competências reconhecidas e a preços de "saldo". Daqui surge também o risco de serem criados cursos sem grande suporte ao nível de recursos humanos com competências, o que poderá levar a desilusões e resultados menos bons. Fica pois o alerta, quer para os agentes educativos, quer para os potenciais estudantes que devem analisar com cuidado as equipas por detrás das ofertas que certamente começarão a abundar.

Do lado da indústria importa que os agentes se apercebam que a coleção de dados com qualidade sobre as suas atividades poderá ser uma mais valia crítica num futuro que se avizinha muito próximo. Quantos mais dados, mais potencial para extração de informação útil, o que poderá levar a ganhos consideráveis no processo de decisão. Importa pois que sejam efetua-

LONGE VÃO OS TEMPOS EM QUE OS DADOS SOBRE A ATIVIDADE DA NOSSA ORGANIZAÇÃO SE RESUMIAM A UMA PEQUENA TABELA NUMA QUALQUER FOLHA DE CÁLCULO. HOJE TEMOS SENSORES EM TEMPO REAL A PRODUZIR INFORMAÇÃO ÚTIL, TEMOS VÍDEOS, TEMOS IMAGENS, TEMOS DOCUMENTOS DE TEXTO, TEMOS SOM, ENFIM TEMOS TODA UMA PROFUSÃO DE FORMATOS, FONTES E TIPOS DE DADOS, QUE COMPLICAM SOBREMANEIRA A TAREFA DE EXTRAIR INFORMAÇÃO ÚTIL DOS MESMOS

dos investimentos consideráveis nestas áreas, mas que se tenha cuidado com os *players* escolhidos para o efeito. Como qualquer fenómeno de crescimento rápido, e com muito dinheiro em jogo, a *Data Science* também está sujeita a muitas falsas promessas e muito *wishful thinking* que frequentemente leva a desperdícios de recursos e resultados aquém do esperado. Muitas vezes, um investimento em equipas dentro de casa, usando uma das muitas ferramentas gratuitas e com grande qualidade que existem disponíveis, acaba por ser mais rentável do que contratos de grandes números com grandes consultoras.

REFERÊNCIAS

[1] L. Torgo; E. Lopes (2011): "Utility-Based Fraud Detection". In: IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Ed. by T. Walsh. AAAI Press: IJCAI/AAAI.

AGREGAÇÃO ÓTIMA PARA GESTÃO DO CONSUMO ELÉTRICO EM REDES INTELIGENTES



MIGUEL F. ANJOS

Polytechnique Montréal, Canadá
miguel-f.anjos@polymtl.ca

A eletricidade é uma forma de energia crítica para a nossa sociedade. As redes elétricas que existem hoje em dia foram construídas baseadas num modelo de produção de eletricidade em grandes quantidades (para tirar partido das economias de escala) acompanhada de uma rede de grande escala para levar essa eletricidade até aos consumidores. Este modelo industrial está a ser transformado pela combinação de vários desenvolvimentos nos últimos anos, principalmente o aumento da produção de eletricidade de fontes eólica e solar, a quantidade quase ilimitada de dados disponíveis através dos contadores inteligentes e outros aparelhos instalados na rede elétrica, o rápido aumento da capacidade das novas tecnologias de armazenamento de energia, e a eletrificação crescente dos meios de transporte. Estes desenvolvimentos levam à criação das *smart grids* (redes inteligentes).

As *smart grids* são a combinação de um sistema tradicional de produção, transmissão e distribuição de energia elétrica com *fluxos bidirecionais de energia e de dados* entre produtores e consumidores. Esta nova configuração apoia significativamente a integração das fontes eólica e solar assim como das novas tecnologias

mencionadas acima. O impacto das *smart grids* conta-se em economias de energia, redução dos custos, e aumento da fiabilidade e da segurança do sistema elétrico.

No processo de transformação das redes elétricas atuais em *smart grids*, é preciso resolver várias questões práticas. Um aspeto central é que qualquer sistema elétrico tem de assegurar que a produção e o consumo de eletricidade estejam continuamente, e quase perfeitamente, em equilíbrio. Os sistemas tradicionais mantêm este equilíbrio ajustando a produção para corresponder ao consumo, mas isso só é possível desde que a produção elétrica de fontes reguláveis seja suficiente para fazer este ajustamento. Com o aumento das contribuições eólica e solar que não são reguláveis, a proporção da produção ajustável diminui. Não obstante o progresso nos meios de armazenamento de energia, esta nova realidade expõe a necessidade de assegurar o equilíbrio ajustando o consumo conforme necessário. Abre-se assim a possibilidade aos consumidores de participar na operação do sistema. Ao conjunto das diferentes abordagens disponíveis para obter esta participação no funcionamento do sistema dá-se o nome de *demand-side management* (gestão do consumo) ou também *demand response*.

A companhia francesa Schneider Electric é uma multinacional especializada na gestão de energia e automação, desenvolvendo tecnologias e soluções integradas para gerir o abastecimento e a utilização da energia de forma segura, fiável, eficiente e sustentável. A Schneider Electric tem uma presença internacional no ramo de *demand response* através da companhia parceira francesa Energy Pool. Foi no quadro desta parceria que se concretizou um projeto de IO em *demand response* em colaboração com a Polytechnique Montréal.

Fundada em 2009, a Energy Pool especializa-se em serviços de agregação e *demand response* para consumidores industriais e comerciais. Um *agregador de energia* é um agente que intervém num mercado de energia através de ofertas de mudança do consumo durante um certo período numa zona específica da rede. Na prática trata-se principalmente de redução de consumos, pois os momentos de alto nível de consumo, chamados *demand peaks* (picos de consumo), são uma grande

preocupação para os operadores do sistema elétrico. Uma mudança do consumo pode, no entanto, também ser um aumento pois é benéfico equilibrar a rede com um consumo superior quando a produção de energia eólica e solar é abundante e a capacidade de armazenagem da energia excedente é insuficiente. A gestão das cargas constitui assim um meio operacional que permite garantir a viabilidade da rede elétrica, particularmente durante os períodos de pico de consumo.

O agregador cumpre os compromissos associados às ofertas de redução da carga através de um portfólio de capacidades de redução sob a forma de contratos com consumidores, que são presentemente, sobretudo empresas industriais e comerciais. Estes consumidores vinculados por contratos com o agregador aceitam, mediante compensação financeira e dentro de parâmetros específicos para cada contrato, reduzir o consumo quando solicitados pelo agregador. O objetivo do agregador é naturalmente de maximizar os seus lucros através da gestão do seu portfólio de contratos com os consumidores, ao quais damos a partir de agora o vocábulo de *recursos*.

As operações do agregador apresentam uma gama de problemas de IO, tanto estratégicos (por exemplo determinar os termos dos contratos) como táticos e operacionais. A colaboração com a Polytechnique Montréal em 2012 estudou os aspetos táticos e operacionais de decidir quais recursos afetar a cada redução de consumo assim que como assegurar capacidade suficiente de substituição dos recursos em caso de imprevistos. Este projeto foi executado entre 2011 e 2013 por uma equipa franco-canadiana de sete investigadores [1].

O esquema de base para a operação do agregador apresenta-se da seguinte forma. O agregador gere dois portfólios de contratos bilaterais que geralmente abrangem vários meses. Um dos portfólios é constituído de contratos com o operador da rede, e o outro de contratos com os recursos que participam na oferta de *demand response*. Um contrato com o operador é definido por quantidades de redução de energia a serem realizadas, mediante solicitação pelo operador da rede, dentro de períodos especificados. Um contrato com um recurso especifica uma redução do consumo de energia, mediante solicitação do agregador, dentro

A ENERGY POOL TORNOU-SE O PRIMEIRO AGREGADOR NA FRANÇA A OFERECER O SERVIÇO DE *FREQUENCY CONTAINMENT RESERVE* QUE É O SERVIÇO DE *DEMAND RESPONSE* MAIS EXIGENTE E, CONSEQUENTEMENTE, O MAIS VALIOSO

dos prazos especificados. Geralmente os contratos com os recursos também impõem máximos quanto ao número de solicitações que o agregador pode executar, assim como a possibilidade do recurso recusar-se a participar, sob certas condições.

Sendo os portfólios de contratos fixos, o agregador responde às solicitações do operador encaminhando-as para um grupo apropriado de recursos cuja ação conjunta resulta na redução de energia desejada. Após pagamento do operador pelo serviço de redução, o agregador faz os pagamentos devidos aos recursos e fica com a diferença. Como o pagamento do operador e o custo das reduções são pré-determinados pelos contratos, o problema tático que nos interessa é a minimização dos custos de mobilização dos recursos, que explicamos de seguida.

No início de cada semana, antes que a rede faça qualquer solicitação, o agregador transmite aos recursos um plano de mobilização que atribui recursos disponíveis e compatíveis às solicitações que possivelmente serão feitas pelo operador. Este plano tem de garantir, *para qualquer combinação possível de solicitações*

do operador, que os recursos mobilizados sejam suficientes para corresponder às mesmas, porque só os recursos mobilizados podem participar nas reduções de consumo durante a semana. O plano de mobilização semanal também toma em conta os limites de disponibilidade de cada recurso ao longo do horizonte de planificação, por exemplo seis meses. O agregador também tem de estabelecer capacidade extra suficiente no caso de certos recursos se recusarem a participar. Como o ato de mobilizar um recurso incorre num custo, o critério de otimização é a minimização dos custos de mobilização.

Este problema de planificação semanal requer simultaneamente uma planificação com meses de antecedência para gerir os limites do número de solicitações para cada recurso. Ambas planificações também envolvem aspetos estocásticos (incerteza quanto às solicitações do operador que se vão concretizar, e quanto à disponibilidade de cada recurso), e de complexidade combinatória (combinações possíveis de recursos para responder a cada solicitação).

O resultado do projeto, publicado em [2], foi uma abordagem que combina dois modelos de otimização matemática. Um modelo é um problema de otimização linear em números inteiros que calcula o plano de mobilização ótimo para uma semana. Como este primeiro modelo não toma em conta o que acontece além da próxima semana, é utilizado a par com outro modelo que utiliza uma perspetiva a longo prazo para assegurar uma gestão ótima do portfólio de capacidades durante o horizonte de planificação. Esta otimização a longo prazo é realizada de maneira eficaz utilizando o que chamamos *configurações de disponibilidade* que permitem considerar as necessidades futuras sem ter de recorrer à modelização por cenários (que seria demasiado pesada para esta aplicação). Os modelos foram concebidos de maneira a que a solução do modelo de longo prazo passe a informação necessária ao

modelo a curto prazo para que o resultado final responda às exigências do agregador. A abordagem consegue assim garantir ao agregador um alto nível de fiabilidade.

O impacto deste projeto foi significativo. Em 2013, a Energy Pool tornou-se o primeiro agregador na França a oferecer o serviço de *Frequency Containment Reserve* (reserva primária na rede elétrica) que é o serviço de *demand response* mais exigente, e consequentemente o mais valioso. Nos finais de 2016, a Energy Pool geria uma capacidade total de mais de 3300 MW em vários países, incluindo a França, a Bélgica, o Reino Unido, a Noruega, e o Camarões, tendo também projetos no Japão, na Polónia e na Turquia. Os seus clientes operam nos ramos do aço, da eletrometalurgia, do papel, do cimento, da saúde, etc. Através da utilização de técnicas de IO, incluindo os algoritmos criados em colaboração com a Polytechnique Montréal, a Energy Pool oferece aos seus clientes uma redução até 15% dos seus custos energéticos. A colaboração entre a Schneider Electric e a Polytechnique Montréal no ramo do *demand response* continua no quadro da nova NSERC-Hydro-Quebec-Schneider Electric Industrial Research Chair inaugurada em maio de 2017 [3].

ATRAVÉS DA UTILIZAÇÃO DE TÉCNICAS DE IO, INCLUINDO OS ALGORITMOS CRIADOS EM COLABORAÇÃO COM A POLYTECHNIQUE MONTRÉAL, A ENERGY POOL OFERECE AOS SEUS CLIENTES UMA REDUÇÃO ATÉ 15% DOS SEUS CUSTOS ENERGÉTICOS

REFERÊNCIAS

- [1] Miguel F. Anjos, Joseph Biron, François Gilbert, Davy Marchand-Maillet, Patrice Marcotte, Stéphanie Petit-Halajda, and Gilles Savard. *Agrégation optimale d'effacements en demand-response*. Apresentação no congresso internacional ROADEF 2012.
- [2] François Gilbert, Miguel F. Anjos, Patrice Marcotte, Gilles Savard. *Optimal design of bilateral contracts for energy procurement*. *European Journal of Operational Research*, Volume 246, Issue 2, 2015, 641-650, <https://doi.org/10.1016/j.ejor.2015.04.050>.
- [3] <http://www.polymtl.ca/carrefour-actualite/nouvelles/inauguration-de-la-chaire-de-recherche-industrielle-crsng-hydro-quebec-schneider-electric-en>



JOÃO N.C.GONÇALVES

Centro I&D Algoritmi
Departamento de Produção e Sstemas,
Universidade do Minho, Braga, Portugal
jncostagonalves@gmail.com



HELENA SOFIA RODRIGUES

Escola Superior de Ciências Empresariais,
Instituto Politécnico de Viana do Castelo,
Valença, Portugal
CIDMA, Departamento de Matemática,
Universidade de Aveiro, Aveiro, Portugal
sofiarodrigues@esce.ipv.pt



M.TERESA T. MONTEIRO

Centro I&D Algoritmi,
Departamento de Produção e Sstemas,
Universidade do Minho, Braga, Portugal
tm@dps.uminho.pt

CONTROLO ÓTIMO NA DIFUSÃO DE MENSAGENS DE MARKETING VIRAL

1. INTRODUÇÃO

O Marketing revela-se uma preciosa ferramenta na potenciação e alavancagem de qualquer ideia de negócio ou empresa. Neste sentido, no seio empresarial, é de suma importância conceber estratégias capazes de colmatar as falhas das campanhas de marketing tradicionais. Uma destas estratégias é comumente conhecida como Marketing Viral (MV), que pode ser vista como o uso do marketing passa-a-palavra (*word-of-mouth*) na replicação e difusão de uma determinada mensagem para um vasto conjunto alvo de consumidores [1]. O MV visa atingir as massas com um baixo custo tirando partido dos efeitos de rede e evita a necessidade de estabelecer um contacto direto com a população alvo.

No entanto, esta estratégia comporta algumas desvantagens como a natureza incontrolável do processo ou ainda a dificuldade em controlar o *timing* e a correta transmissão da mensagem [2]. É precisamente neste sentido que tem lugar a Teoria do Controlo Ótimo (TCO), ao utilizar funções de controlo que permitem otimizar estratégias de marketing no sentido de maximizar a difusão da informação sem comprometer a capacidade financeira da empresa. No que concerne à aplicação da TCO ao MV, recomenda-se a leitura de [3] e das referências aí citadas.

Um aspeto interessante e fulcral na construção do modelo matemático subjacente a este artigo prende-se com o facto de que as dinâmicas de marketing podem, efetivamente, ser modeladas por fenómenos relacionados com Epidemiologia Matemática, vide, e.g., [1] – estabelecendo fortes analogias com a propagação de um vírus numa determinada população. O presente artigo, publicado em [4], estuda não só a dinâmica e impacto de uma das maiores publicidades virais de 2013 – *Dove Real Beauty Sketches*, mas também as janelas temporais ótimas para aplicação de algumas estratégias de marketing. Neste sentido, atendendo ao facto de que uma das maiores dificuldades em difundir mensagens de sucesso virais se relaciona com a criação de mecanismos que consigam persuadir um indivíduo a disseminar a mensagem nas janelas temporais ótimas, é explorada a incorporação da TCO num modelo epidemiológico SIR (Suscetíveis-Infetados-Recuperados). São comparados e simulados di-

ferentes cenários de custo, de modo a fornecer algumas linhas de orientação a profissionais de Marketing.

2. MODELO EPIDEMIOLÓGICO SIR

O modelo SIR é um modelo epidemiológico que divide a população em três compartimentos mutuamente exclusivos: *S* (Suscetíveis), *I* (Infetados) e *R* (Recuperados). É, no entanto, possível reescrever o significado destes compartimentos num contexto de marketing em detrimento de um sentido epidemiológico, i.e.:

S: Conjunto de indivíduos suscetíveis a difundir a mensagem de marketing;

I: Conjunto de indivíduos que difundem a mensagem de marketing;

R: Conjunto de indivíduos que param de difundir a mensagem de marketing, uma vez que, por exemplo, a mensagem deixa de ser atrativa.

A dinâmica deste modelo ao longo do tempo pode ser descrita pelo seguinte sistema de equações diferenciais ordinárias com condições iniciais:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N}, & S(0) &= S_0 > 0 \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \gamma I(t), & I(0) &= I_0 > 0 \\ \frac{dR(t)}{dt} &= \gamma I(t), & R(0) &= 0. \end{aligned} \quad (1)$$

A população total é dada por $N = S(t) + I(t) + R(t)$. Desta forma, o sistema de equações (1) pode ser reduzido para

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N}, & S(0) &= S_0 > 0 \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \gamma I(t), & I(0) &= I_0 > 0. \end{aligned} \quad (2)$$

Explorando a dinâmica do sistema de equações (2), quando a mensagem de marketing apresenta características que motivem o indivíduo a difundir-la, os indivíduos da classe *S* passam a integrar a classe *I* a uma taxa β . Quando, por sua vez, os indivíduos deixam de difundir a mensagem de marketing, passam a integrar a classe *R* a uma taxa γ .

Com base nos primeiros 7 dias da campanha e estimando a população total como $N=10^9$ (número aproximado de YouTuber's em 2013), os parâmetros β e γ foram estimados recorrendo à rotina `FMINSEARCH` presente na *toolbox* de otimização do Matlab,

baseada no algoritmo Nelder-Mead. A escolha desta janela temporal prende-se diretamente com a obtenção do pico máximo de difusão da campanha.

Estabelecido o suporte teórico sobre o qual o modelo epidemiológico assenta, procurou-se estudar soluções de controlo que permitissem, por um lado, maximizar a difusão da campanha e, por outro lado, minimizar os custos associados. Para isso recorreu-se à formulação de um problema de controlo ótimo, especificado na secção seguinte.

3. PROBLEMA DE CONTROLO ÓTIMO

Como proposto em [3], e após a normalização do sistema de equações (2), foi formulado um problema de controlo ótimo resultante da incorporação de duas funções de controlo, $u_1(t)$ e $u_2(t)$, no modelo. Assim, lembrando que $r(t) = 1 - s(t) - i(t)$, o sistema de equações ordinárias com controlo ótimo resultante é dado por:

$$\frac{ds(t)}{dt} = -[\beta + u_2(t)]s(t)i(t) - u_1(t)s(t), \quad s(0) = \frac{S(0)}{N}$$

$$\frac{di(t)}{dt} = [\beta + u_2(t)]s(t)i(t) + u_1(t)s(t) - \gamma i(t),$$

$$i(0) = \frac{I(0)}{N}. \quad (3)$$

As funções de controlo $u_1(t)$ e $u_2(t)$ representam, respetivamente, o recrutamento de indivíduos suscetíveis para atuar como *spreaders* sob a população alvo e o incentivo de indivíduos previamente 'infetados' para continuar a difundir a mensagem – mantendo níveis significativos de epidemia por um maior período de tempo. Repare-se, portanto, que o principal objetivo em marketing é gerar uma epidemia em detrimento de mitigá-la, como no caso tradicional da epidemiologia.

Seja t_f o tempo final considerado para a difusão da campanha em questão. Define-se o conjunto de funções de controlo admissíveis como sendo

$$\Omega = \{(u_1(\cdot), u_2(\cdot)) \in (L^2(0, t_f))^2 \mid 0 \leq u_1(t) \leq u_{1max}$$

$$\wedge 0 \leq u_2(t) \leq u_{2max} \quad \forall t \in [0, t_f]\}.$$

Para $u_i, i \in \{1, 2\}$, a aplicação do controlo é máxima se $u_i = u_{imax}$ e mínima se $u_i = 0$. Assim, o principal objetivo prende-se com encontrar os valores ótimos u_1^* e u_2^* para os controlos u_1 e u_2 , de tal maneira que as trajetórias s e i sejam a solução do sistema (3) ao longo do intervalo $[0, t_f]$, e maximizar o funcional

$$\max_{\Omega} J((u_1(\cdot), u_2(\cdot))) = i(t_f) + r(t_f) + \int_0^{t_f} -[B u_1^2(t) + C u_2^2(t)] dt, \quad (4)$$

sujeito a (3), onde as constantes não negativas B e C representam os pesos relativos ao custo do investimento nos controlos u_1 e u_2 , respetivamente. Neste contexto, o funcional pode ser visto como um *trade-off* entre a maximização da difusão da informação no fim da campanha e a minimização dos custos a ela associados.

4. SIMULAÇÕES NUMÉRICAS E RESULTADOS

Nesta secção, a influência da incorporação das variáveis de controlo no modelo é avaliada através de simulações em Matlab com recurso ao algoritmo *Forward-Backward Sweep* (vide, e.g., [5]). Após calcular o funcional (4) para uma série de cenários de custo de investimento em estratégias de controlo, são agora apresentados os casos que induzem um maior funcional ($(B, C) = (1, 10^{-3})$ - baixo custo de investimento e $(B, C) = (1, 10)$ - alto custo de investimento). Importa ressaltar que não

AS DINÂMICAS DE MARKETING PODEM, EFETIVAMENTE, SER MODELADAS POR FENÓMENOS RELACIONADOS COM EPIDEMIOLOGIA MATEMÁTICA, ESTABELECIDO FORTES ANALOGIAS COM A PROPAGAÇÃO DE UM VÍRUS NUMA DETERMINADA POPULAÇÃO

foram considerados casos em que $B < 1$ dado que a existência de uma epidemia de marketing passa por um investimento inicial em recrutar indivíduos suscetíveis para difundir a mensagem. Como exemplos de cenários de baixo e alto custo de investimento em estratégias de marketing

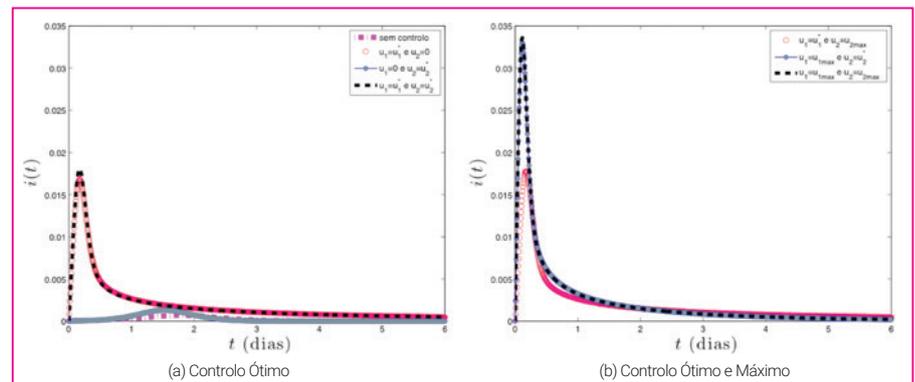


Fig. 1 - $i(t)$ com $(B, C) = (1, 10^{-3})$ e $u_{1max} = 0.4, u_{2max} = 1$.

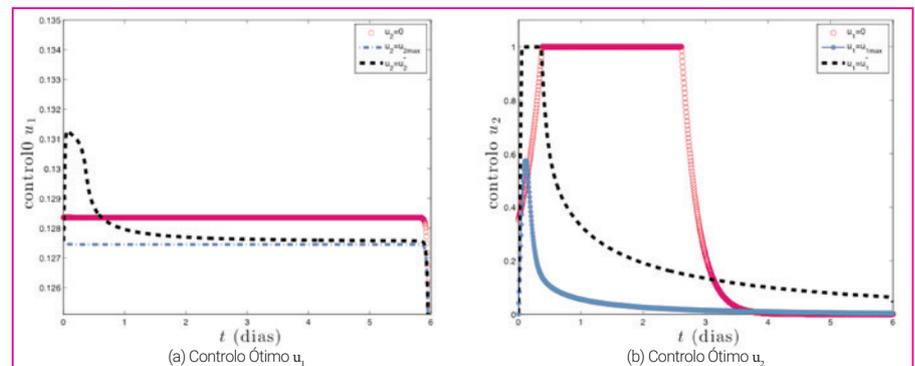


Fig. 2 - Controlos ótimos u_1 e u_2 com $(B, C) = (1, 10^{-3})$ e $u_{1max} = 0.4, u_{2max} = 1$.

OS NÍVEIS DE DIFUSÃO DE INFORMAÇÃO AUMENTAM A TAXAS CONSIDERAVELMENTE MAIORES QUANDO A TCO É APLICADA. MAIS AINDA, CONCLUI-SE QUE QUANTO MAIOR É O CUSTO DE INVESTIMENTO EM POLÍTICAS DE CONTROLO DE MARKETING, MENORES SÃO OS NÍVEIS DE DIFUSÃO DE INFORMAÇÃO

destacam-se, por exemplo, a exploração de redes sociais e a oferta de recompensas monetárias, respetivamente.

4.1. Cenário de custos de investimento baixos

Pela Fig.1(a), é possível constatar que o número de indivíduos infetados é sempre maior com a aplicação de políticas de controlo do que sem qualquer aplicação. Mais ainda, a análise das Figs. 2(a) e 2(b) mostra que a combinação de controlos (u_1^*, u_2^*) é a melhor estratégia para atingir as massas num curto espaço de tempo. Para $t \in [1, 2]$, o controlo u_2 deve ser aplicado por forma a cativar indivíduos infetados a continuar a difundir a mensagem. Note-se que apesar de os níveis de infeção serem maiores quando a combinação de políticas de controlo ótimo e máximo são aplicadas, resultaram valores de funcional quase nulos - o que mostra que a aplicação exacerbada de estratégias de controlo pode comprometer a capacidade financeira da empresa, não cumprindo o objetivo de minimização de custos.

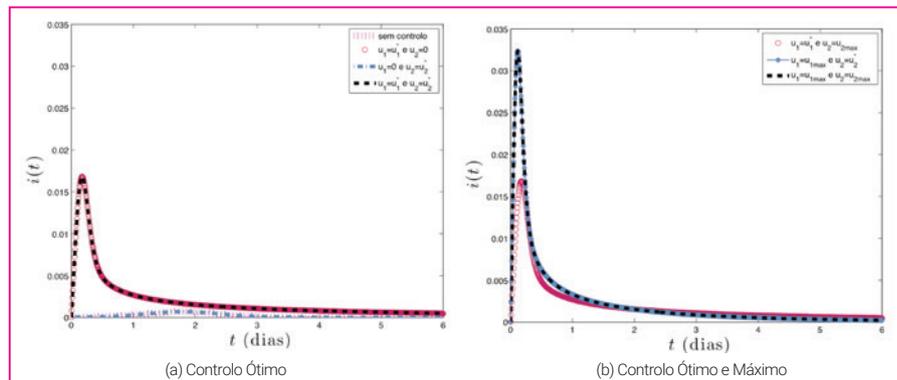


Fig. 3 - $i(t)$ com $(B,C)=(1,10)$ e $u_{1max}=0.4$, $u_{2max}=0.01$.

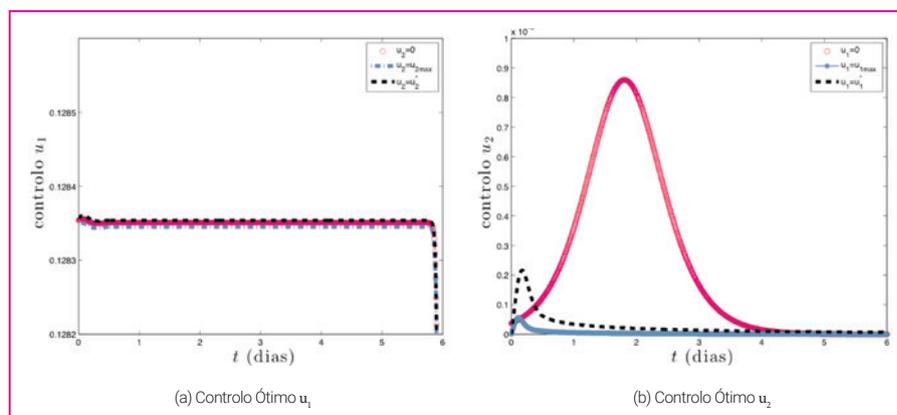


Fig. 4 - $i(t)$ com $(B,C)=(1,10)$ e $u_{1max}=0.4$, $u_{2max}=0.01$.

4.2 Cenário de custos de investimento altos

Comparando com o cenário anterior, o aumento do peso de custo faz com que a magnitude do controlo seja residual, permitindo inferir que um maior investimento em estratégias de controlo não induz maiores níveis de infeção (ver Figs. 3 e 4).

5. CONCLUSÕES

Como principal conclusão deste estudo, destaca-se que os níveis de difusão de informação

aumentam a taxas consideravelmente maiores quando a TCO é aplicada. Mais ainda, conclui-se que quanto maior é o custo de investimento em políticas de controlo de marketing menores são os níveis de difusão de informação $(i(t))$. Finalmente, como investigação futura, impõe-se a necessidade de oferecer maior robustez ao modelo apresentado, fazendo com que contemple um maior número de variáveis intrínsecas à real natureza do marketing viral, dado o seu carácter errático.

REFERÊNCIAS

- [1] Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. *ACM Trans. Web (TWEB)* 1(1), 1-39 (2007)
- [2] Woerndl, M., Papagiannidis, S., Bourlakis, M., Li, F.: Internet-induced marketing techniques: Critical factors in viral marketing campaigns. *Int. Journal of Business Science and Applied Management* 3(1), 33-45 (2008)
- [3] Kandhway, K., Kuri, J.: How to run a campaign: Optimal control of SIS and SIR information epidemics. *Applied Mathematics and Computation* 231, 79-92 (2014)
- [4] Gonçalves, J.N.C., Rodrigues, H.S., Monteiro, M.T.T.: Optimal control strategies for an advertisement viral diffusion. *XVIII Congress of Operational Research. Springer Proceedings in Mathematics and Statistics*. In press (2017)
- [5] Lenhart, S., Workman, J.: *Optimal Control Applied to Biological Models*, Chapman & Hall/CRC, Boca Raton, FL (2007)

O SÓCIO N.º ... 258

Depois de obter o curso de Ciências Matemáticas na Universidade de Coimbra em 1967 tive de uma curta experiência como professora, em diversas disciplinas de Ciências, no liceu de Moçâmedes, em Angola, em 1967/1968. Entrei como assistente eventual na Faculdade de Ciências Universidade do Porto em outubro de 1968. A UP foi a primeira universidade portuguesa a ter um computador para ensino dos alunos de diversos cursos, a partir de 1967. Comecei com Análise Numérica e estive um ano a aprender e a ensinar, em aulas práticas desta disciplina, regida pelo Professor Rogério Nunes. No fim desse ano, passei à regência teórica dessa matéria. O Professor Rogério Nunes por volta de 1969 levou-me à biblioteca, mostrou-me uma secção com livros de Matemática Aplicada e disse-me para escolher um assunto e que o estudasse. Folhiei diversos livros e interessou-me um, Linear Programming and Extensions, do Georges Dantzig. Estudei a matéria e fui falar com o Professor que me disse para preparar um curso com base nesses conhecimentos; foi assim que regii uma disciplina de Programação Linear e Extensões no ano letivo de 1971/1972, nova na nossa Faculdade. A partir daí fiquei interessada em Optimização e suas aplicações. Tive a oportunidade de fazer um estágio no Institut de Programmation, em Paris, no verão de 1971. Em 1972-73 obtive o grau de Master of Arts em IO na Escola de Business da Universidade de Lancaster. O meu trabalho de mestrado, de carácter muito aplicado, foi Competitive Bidding in Contract Furniture Industry. O que de mais importante aprendi em Lancaster foi que a IO era um modo de resolver problemas do mundo real: se a situação fosse quantificável e houvesse dados, construir um modelo matemático e resolve-lo, discutindo os resultados perante a realidade; noutras situações usar o bom senso e a inteligência.

De regresso a Portugal ocupei-me com a regência teórica e prática de diversas disciplinas na área de IO, o que me permitiu aprofundar os aspetos matemáticos e aplicados de todas as técnicas de IO: optimização (linear, não linear, inteira, estocástica, robusta), programação dinâmica, grafos, redes, processos estocásticos com ênfase em filas de espera, teoria de jogos, simulação. Escrevi, em FORTRAN IV, programas de optimização de funções de diversas

variáveis; interessei-me por Programação Inteira e escrevi algumas notas sobre o assunto. Já como assistente convidada e de certo modo pressionada a fazer um doutoramento, regresssei a Inglaterra em 1983, onde na School for Industrial and Business Studies, da Universidade de Warwick, me doutorei em IO com uma tese com o título Mathematical Models in Capital Investment Appraisal, tendo obtido conhecimentos nas áreas de economia e finanças. Além de disciplinas na área de Probabilidades e Estatística, e de Métodos Numéricos, mantive-me sempre ligada, em termos letivos, a IO e fui orientadora de diversos estágios curriculares em empresas muito diversificadas. Estes estágios eram do âmbito de licenciatura (Matemática Aplicada à Tecnologia) e de mestrado (MAOPI – Modelação, Análise e Optimização de Processos Industriais; Engenharia Matemática). Fui membro da SPEIO (Sociedade Portuguesa de Estatística e Investigação Operacional), onde a IO era o parente pobre: só mais tarde se organizou a Sociedade Portuguesa de Estatística (SPE) e também a APDIO. Ainda sou membro destas duas sociedades. Muito mudou no nosso país, ao longo destes anos, relativamente à IO embora me pareça que ainda se pensa muito nesta área como sendo apenas optimização e de forma bastante académica.



MARIA DO CARMO GUEDES

Faculdade de Ciências, Universidade do Porto
mmguedes@fc.up.pt

O QUE DE MAIS IMPORTANTE APRENDI EM LANCASTER FOI QUE A IO ERA UM MODO DE RESOLVER PROBLEMAS DO MUNDO REAL: SE A SITUAÇÃO FOSSE QUANTIFICÁVEL E HOUVESSE DADOS, CONSTRUIR UM MODELO MATEMÁTICO E RESOLVÊ-LO, DISCUTINDO OS RESULTADOS PERANTE A REALIDADE; NOOUTRAS SITUAÇÕES, USAR O BOM SENSO E A INTELIGÊNCIA

IO 2017



Sessão de Abertura do Congresso IO 2017 (da esq. para a dir.: Lia Oliveira (Chair da Comissão Organizadora), Ismael Vaz (Chair da Comissão Científica), José Fernando Oliveira (Presidente da Comissão Diretiva da APDIO), Florbela Correia (Vice-Presidente do Instituto Politécnico de Viana do Castelo), Jorge Mendes (Presidente da Câmara Municipal de Valença))

IO 2017: Investigação Operacional Solidária

Valença foi a anfitriã do XVIII Congresso da Associação Portuguesa de Investigação Operacional (APDIO) que decorreu entre 28 e 30 de junho de 2017 nas instalações da Escola Superior de Ciências Empresariais do Instituto Politécnico de Viana do Castelo.

À semelhança das edições anteriores, o IO2017 reuniu profissionais, investigadores, docentes e estudantes interessados em Investigação Operacional com o objetivo de aprender, trocar histórias de sucessos e insucessos, conviver e iniciar novas colaborações.

Procurou-se centrar a ideia em torno da "Investigação Operacional Solidária" e por isso associámo-nos a uma causa com a qual todos nos identificamos - A operação Nariz Vermelho. Demos voz aos seus protagonistas para que partilhassem o que fazem e nos abrissem a porta para possíveis parcerias de colaboração face aos seus problemas, dando-nos a oportunidade de ajudar cientificamente. O nosso painel de oradores plenários foi constituído por Ana Póvoa (Instituto Superior Técnico da Universidade de Lisboa), Ruth Kaufman (The OR Society) e Ricardo Vargas (Brightline Initiative), que superou todas as expectativas e contemplou-nos com as suas experiências na liderança de

projetos associados aos mais variados projetos na temática do congresso. Em paralelo, existiu um esforço particular em promover a cidade de Valença, que teve o seu lugar de destaque através do programa social cuidadosamente desenvolvido com o apoio da Câmara Municipal. Não podíamos deixar os participantes partirem desconhecendo a história escondida em cada pedra das muralhas e sem usufruírem da beleza natural que a cidade oferece. A gastronomia da região também foi devidamente representada, e de certo obteve admiradores.

O congresso mudou-se novamente para Norte, e superou as expectativas. Mais de 150 participantes, provenientes de mais de 30 instituições localizadas em 7 países, para participarem nestes três dias dedicados à partilha de conhecimento.

A comissão organizadora do IO2017 agradece à APDIO a confiança depositada ao possibilitar-nos esta experiência. Agradecemos a todos os nossos parceiros/apoios, CIDMA, SISCOG e em particular ao IPVC e à Câmara Municipal de Valença que nos facultou apoios fundamentais na logística do congresso e reconheceram o valor deste tipo de eventos.

A todos os presentes, Obrigada pela vossa colaboração porque afinal sem vocês não teria sido possível e até ao IO2018 na bela cidade de Aveiro!

IO 2018: Criar valor para a sociedade

O IO2018 - XIX Congresso da Associação Portuguesa de Investigação Operacional realizar-se-á na Universidade de Aveiro, entre 5 e 7 de setembro de 2018. A abordagem de problemas complexos, muitos deles com origem em parcerias entre diferentes organizações, usando metodologias, modelos e técnicas da investigação operacional, permite contribuir para a melhoria da eficiência, competitividade e sustentabilidade das organizações e da sociedade de uma forma geral.

No IO2018 em particular, pretende-se realçar o papel que a IO tem na criação de valor para diferentes setores da sociedade e na forma como poderá, no futuro, aumentar o impacto da sua intervenção, assim como atuar sobre novas áreas.

Assim, apelamos à vossa participação através da submissão de trabalhos que podem ser apresentados quer na língua portuguesa quer na inglesa, podendo encontrar mais informação no site <http://apdio.pt/web/io2018>.

Esperamos-vos em Aveiro no IO2018!



Prémios Obtidos

José Rui Figueira,

Professor Associado com Agregação no Departamento de Engenharia e Gestão do Instituto Superior Técnico – Universidade de Lisboa, recebeu a Gold Medal da International Society on Multiple Criteria Decision-Making em julho de 2017. Esta é a maior distinção que a International Society on Multiple Criteria Decision Making concede a quem, ao longo de uma carreira distinta, contribuiu significativamente para a teoria, metodologia, prática e desenvolvimento profissional do MCDM.



Entrega da Gold Medal a José Rui Figueira durante a 24th International Conference on Multiple Criteria Decision Making, julho 2017, Ottawa, Canadá (da esq. para a dir: Theodor Stewart (President of the Award Committee), José Rui Figueira, Sarah Ben Amor (Organiser of the MCDM 2017 Conference, Ottawa, Canada) e Murat Koksalan (President of the International Society on MCDM)).

Miguel F. Anjos,

Professor Catedrático no Department of Mathematics and Industrial Engineering of Polytechnique Montréal, recebeu o título de "EUROPT fellow" em julho de 2017. A EUROPT Fellow é escolhida anualmente pela Comissão Diretiva da EUROPT – The Continuous Optimization Working Group of EURO, com base nas nomeações recebidas dos membros da EUROPT.



Entrega do Título "EUROPT Fellow" a Miguel F. Anjos (à esq.) pelo Presidente da Comissão Diretiva da EUROPT Julius Žilinskas (à dir.), durante o 15th EUROPT Workshop on Advances in Continuous Optimization, julho 2017, Montréal, Canadá

Teses de Doutoramento

Autor: Miguel Jorge Vieira

Título: Towards the Development of a Decision-support Tool for the Production Planning and Campaign Scheduling of Biopharmaceutical Facilities
Instituição: Instituto Superior Técnico, Universidade de Lisboa / MIT Portugal Program

Programa de Doutoramento: Líderes para Indústrias Tecnológicas (EDAM-LTI)

Data de conclusão: Outubro de 2017

Orientador: Tânia Pinto Varela

Co-Orientador: Ana Paula Barbosa Póvoa

Prémio APDIO

Pelo oitavo ano consecutivo, foi atribuído o Prémio APDIO - FCT NOVA ao melhor aluno de Investigação Operacional da Faculdade de Ciências e Tecnologia da UNL. Este Prémio foi constituído com parte dos lucros do IO2009 - 14º congresso nacional da APDIO, que decorreu na FCT NOVA.

O Prémio relativo ao ano letivo de 2016/17 foi atribuído a Raquel Duarte Macedo do Mestrado Integrado em Engenharia Informática. O Diploma e cheque correspondentes ao Prémio foram entregues na Sessão Comemorativa do 40º Aniversário da FCT NOVA, em novembro 2017.

BERNARD ROY

1934-2017

Brief description of achievements of Bernard Roy for the international community of Operational Research

(by: Roman Słowiński, José Rui Figueira, and Salvatore Greco, Coordinators of the EURO WG on MCDA)



Professor Bernard Roy was a pioneer of Operational Research in Europe. He was a world-wide recognized expert in the area of Multiple Criteria Decision Aiding (MCDA). He was a founder of the EURO Working Group on MCDA, and, more generally, a founder of the European School of MCDA. He was also president of EURO and member of the Editorial Board of the European Journal of Operational Research since its conception.

Bernard Roy has made a very significant and innovative contribution to Operational Research. His major breakthrough works were first in graph theory and project scheduling, and then in multiple criteria decision aiding. Among his many achievements, he is the father of the "activity on node" project scheduling technique and of the famous ELECTRE methods. Through his research, teaching, consulting, and service to the community, he was one of the major promoters of OR techniques in Europe.

Bernard Roy was the founder (1974) and Honorary Director of LAMSADE (since 1999), a research group centered on decision aiding, and Président d'honneur of ROADEF (the French association of operational research and decision aiding). He founded in 1975 the EURO Working Group "Multicriteria Decision Aiding" which regularly held two annual meetings since then. He was president of this EURO Working Group during 25 years, and in 2010 he became its honorary chairman. He was President Elect and President of EURO in years 1983-1986.

Bernard Roy held a PhD in Mathematics from the University of Paris (1961); was a graduate of the Statistics Institute of Paris University (1957); and attended courses at Institut d'Etudes Politiques (1954-1956). Before joining the university, from 1957 to 1966, he was a consultant at SEMA (Société d'Economie et de Mathématiques Appliquées), applying Operational Research to real problems. From 1966 to 1972, he was SEMA Scientific Director (METRA International). He joined University Paris-Dauphine in 1972 and created LAMSADE in 1976. From 1979 to 2009, he was RATP (the Paris mass transit authority) Scientific Advisor. Bernard Roy was Doctor Honoris Causa of seven distinguished universities. He received the Prix HERMES de la recherche de la Faculté des Sciences de l'Administration de l'Université Laval (Québec, Canada) in 1991, the EURO Gold Medal in 1992, and the MCDM Gold Medal granted by the International MCDM Society in 1995. In 2015, he was also decorated by the EURO Distinguished Service Medal Award. In September 2017, during the 86th meeting of the EURO Working Group on MCDA in Paris, Bernard Roy received the Medal of the Polish Academy of Sciences corresponding to the highest scientific distinction conferred on personalities exterior to the Academy. He is the author of several books and hundreds of research papers. Bernard Roy has been the advisor of numerous graduate and doctoral students.

Bernard Roy's main contributions focus on two broad themes:

Graph Theory with path-breaking contributions to the theory of flows in networks and project scheduling (with the invention of the activity on nodes method).

Multiple Criteria Decision Aiding with the invention of the family of ELECTRE methods and methodological contributions to decision aiding that have led to the creation of the European School of MCDA.

His recent research addressed two main topics:

Robustness in Operational Research and Decision Aiding.

Multiple criteria tools for collective decision (called concertation in French).

Despite his overwhelming scientific esteem, Professor Bernard Roy was a modest and considerate person. He radiated enthusiasm and kind encouragement for the young. We will keep a grateful memory of him alive.

BOLETIM
APDIO



Associação Portuguesa de Investigação Operacional
Cesur - Instituto Superior Técnico
Av. Rovisco Pais 1049 - 001 Lisboa
T. 218 407 455
apdio@civil.ist.utl.pt

Equipa Editorial

Ana Carvalho
anacarvalho@tecnico.ulisboa.pt
Tânia Ramos
tania.p.ramos@tecnico.ulisboa.pt

Design

Inês Assis
inesassisdesign@gmail.com

Impressão

Gráfica Jorge Fernandes Lda
Tiragem 400 exemplares