BOLETIM APDIO 51

2º Semestre de 2014

Editores: Ana Luísa Custódio Isabel Correia

Apdio 🕁

EDITORIAL

2015 é o ano de mais uma edição do Congresso Nacional da APDIO - IO2015, contando desta vez com o Instituto Politécnico de Portalegre como escola anfitriã. Este encontro, a ter lugar entre 7 e 9 de Setembro de 2015, terá como temática subjacente a "IO e Big Data". Por se tratar de um assunto novo, relevante para a comunidade científica de IO, e que acreditamos não ser do domínio da generalidade dos nossos leitores, procurámos focar o presente número do Boletim nesta temática.

Na verdade, como preparação para o IO2015 e dando continuidade a uma política de aposta na formação contínua, a APDIO foi um dos parceiros na organização de um conjunto de workshops subordinados ao mesmo tema, tal como nos é relatado por João Miranda, no último texto da secção de Notícias. Esta secção conta ainda com um pequeno resumo relativo aos trabalhos de outros dois eventos relevantes para a nossa comunidade: a conferência Optimization 2014 e o encontro anual do EURO Working Group on Operational Research Applied to Health Services (ORAHS 2014). No primeiro caso, o resumo é da autoria de José Valério de Carvalho e Ismael Vaz. No segundo caso, surge pela mão de Maria Eugénia Captivo, Inês Marques e Margarida Vaz Pato.

José Pinto Paixão deixa a sua opinião sobre o papel que a IO e os profissionais de IO poderão desempenhar no actual cenário em que é necessário lidar com um grande volume de informação. Fica a pergunta. Será que estamos preparados para sermos agentes da mudança?

Filipe Carvalho, CEO da Wide Scope, é o entrevistado da actual edição. Procurámos conhecer um pouco melhor esta empresa, nomeadamente no que respeita à sua experiência na utilização de Big Data em aplicações.

Na secção Técnicas de IO, Rodrigo Rodrigues fala-nos do começo do Big Data, num interessante artigo sobre a origem desta área científica. Os conceitos básicos e alguma da terminologia de Data Science são introduzidos por Luís Cavique, na secção IO em Acção.

Tal como vem sendo habitual, na rubrica Portugueses em IO pelo Mundo, procuramos dar a conhecer o percurso de um profissional de IO, português, que exerça a sua actividade no estrangeiro. Nesta edição, o nosso convidado é Miguel Anjos, a trabalhar no Canadá, que, entre outros assuntos, nos alerta para a relevância das aplicações.

A secção Lugar aos Novos foi inaugurada no Boletim 50, onde procuramos divulgar o trabalho desenvolvido pelos nossos colegas mais jovens. Nesta edição, Tatiana Gonçalves de Sousa e Maria Cândida Mourão apresentam-nos um caso de estudo na área da distribuição de gás, integrado na actividade da Galp Energia.

Esperando que este Boletim seja do vosso agrado, desejamos a todos os leitores um excelente 2015! Que saibam aproveitar a disponibilidade de informação, não esquecendo a respectiva qualidade.

Ana Luísa Custódio Isabel Correia





OPTIMIZATION 2014

José Valério de Carvalho, Ismael Vaz, Departamento de Produção e Sistemas, Universidade do Minho

Teve lugar, entre os dias 28 e 30 de julho de 2014, a oitava edição da série de conferências *Optimization*, sob a égide da APDIO, cuja organização esteve a cargo de uma equipa do Departamento de Produção e Sistemas (DPS) da Universidade do Minho, liderada pelos colegas José Valério de Carvalho e Ismael Vaz. O congresso decorreu nas instalações da Escola de Engenharia da Universidade do Minho, Campus de Azurém, em Guimarães.

A sessão de abertura contou com a presença do Vice Presidente da Câmara Municipal de Guimarães, Dr. Amadeu Portilha, da Diretora do DPS, profª. Senhorinha Teixeira, do Presidente da Comissão de Programa, prof. Luís Nunes Vicente, do Presidente da Escola de Engenharia, prof. João Monteiro, dos Co-Presidentes da Comissão Organizadora, prof. José Valério de Carvalho e prof. Ismael Vaz, e do Vice Presidente da APDIO, prof. Carlos Henggeler Antunes (na figura, da esquerda para a direita, respetivamente).

O Optimization 2014 surgiu no seguimento das sete edições anteriores, Optimization 91 (UC, Coimbra), Optimization 95 (UMinho, Braga), Optimization 98 (UC, Coimbra), Optimization 2001 (UA, Aveiro), Optimization 2004 (FCUL, Lisboa), Optimization 2007 (FEP, Porto) e Optimization 2011 (FCT-UNL, Caparica/Lisboa), uma série de conferências lançada pelo prof. Joaquim

João Júdice e que recentemente tem contado com a coordenação do prof. Luís Nunes Vicente.

Os pontos mais altos de um congresso são, naturalmente, as sessões plenárias convidadas. Neste caso foram seis os oradores estrangeiros que enriqueceram o programa científico, cobrindo uma variedade de tópicos, a saber: Masao Fukushima (Some results in multi-leader-follower games), Serge Gratton (Optimization methods for large scale inverse problems), Tim Kelley (Implicit filtering and hidden constraints), Nenad Mladenovic (Clustering community networks by variable neighbourhood search), François Vanderbeck (Lagrangian duality and column generation: A review of solution methods, their common base and specific features, and extensions of the paradigm) e Maarten H. Van der Vlerk (Simple recourse and relatives).

O congresso juntou investigadores nacionais e internacionais que desenvolvem trabalho nas mais diversas áreas científicas da Otimização. A organização contou com 11 sessões organizadas e 21 contribuídas, num total de 32 sessões, distribuídas pelos três dias de congresso e não excedendo mais que 5 sessões em paralelo. Foram proferidas no total 110 comunicações, das quais 74 foram contribuídas, com a participação de oradores de mais de 10 países diferentes. Note-se ainda que a conferência sofreu apenas dois

cancelamentos de última hora, permitindo que praticamente todo o programa científico fosse cumprido conforme o planeado. O empenho e a capacidade dos oradores e dos participantes nacionais mostraram uma comunidade portuguesa de Otimização dinâmica com uma forte ligação internacional.

O programa social contou com uma *welcome reception* no primeiro dia da conferência, um jantar de congresso no segundo dia e um *tour* no terceiro dia, com visita à cidade de Braga, Santuário do Bom Jesus do Monte e Santuário do Sameiro.

Na sessão de encerramento, antes da partida para o tour, foi anunciado o *Optimization 2017*, que decorrerá na Faculdade de Ciências da Universidade de Lisboa, em setembro de 2017, organizado por uma equipa liderada pelos colegas Miguel Constantino e Pedro Moura, e que contará com um evento de comemoração dos 60 anos do prof. Luís Gouveia.



ORAHS 2014

O ORAHS 2014 - 40° encontro anual do EURO Working Group on Operational Research Applied to Health Services (ORAHS) - realizou-se na Faculdade de Ciências da Universidade de Lisboa, entre 20 e 25 de julho de 2014. O grupo encontra-se anualmente, durante uma semana, e conta com um intenso programa científico complementado com contactos com profissionais de saúde. É ainda tradição destes encontros haver um programa social, que proporciona boas oportunidades para discussões informais e para convívio. A conferência deste ano contou com 143 participantes, provenientes de 24 países. Na Tabela 1 fica o registo sobre a distribuição do número de participantes, por país de proveniência. Quanto a Portugal, reuniram-se académicos e profissionais da área, tendo colaborado 10 alunos que prestaram apoio à conferência.

O tema escolhido para a conferência deste ano foi "For Better Practices in Health Care Management", tendo o programa contado com duas comunicações plenárias. Geoff Royston, anterior Presidente da Operational Research Society, falou sobre novas formas de aplicação de investigação operacional aos serviços de saúde, incorporando, nomeadamente, áreas como a análise de decisão, modelos comportamentais e conceção de sistemas ("Exploring New Worlds for Applying Operational Research in Health Services"). Manuel Delgado, Director-Geral da IASIST Portugal, discutiu a utilização de ferramentas de benchmarking para avaliar o desempenho hospitalar, e especificou a metodologia e os indicadores usados pela IASIST na avaliação do desempenho de hospitais portugueses (IAMETRICS). Houve também um debate dedicado ao tema da conferência, centrado na realidade portuguesa, sendo o painel constituído por académicos e profissionais de saúde. Carlos Moreira (Comissão de Avaliação de Medicamentos e INFARMED) moderou o debate, que contou com a participação de Luís Cunha Ribeiro (Presidente da Administração Regional de Saúde de Lisboa e Vale do Tejo), Rui Portugal (Direção-Geral da Saúde, em representação do respectivo Diretor) e João

Carvalho das Neves (Presidente da Administração Central do Sistema de Saúde).

Um dos objetivos dos encontros anuais do grupo ORAHS é promover a relação entre académicos e profissionais de saúde locais. Para isso, é habitualmente incluída, no programa científico destes encontros, uma visita a uma unidade de cuidados de saúde. Este ano, a atividade consistiu numa visita ao Hospital de Santa Maria, onde os participantes foram recebidos por Carlos Martins, Presidente do Centro Hospitalar Lisboa Norte. Na Aula Magna da Faculdade de Medicina da Universidade de Lisboa, Carlos Martins fez uma apresentação do Hospital de Santa Maria, na qual sublinhou as principais valências deste hospital e a atividade dos seus centros tecnológicos inovadores.

O programa incluiu 24 sessões paralelas, onde foram apresentadas 94 comunicações orais, e uma sessão com a apresentação de 7 posters. As sessões abrangeram temas em logística e planeamento de serviços (Operating room planning and scheduling, Workforce planning, Patient scheduling, Patient flow, Ambulance location); economia e políticas de saúde (Health economics, Performance evaluation, Healthcare planning, Regional health service, Risk management); modelação (OR methodologies for home care, Modeling blood delivery, Modeling infectious diseases, Modeling in health care, Disease modeling and policy); e aplicações médicas (Cancer treatment planning).

O programa científico concluiu com a habitual reunião de trabalho com os membros do ORAHS, onde foi feita uma avaliação positiva da conferência. Neste encontro, foi aceite a proposta de organização do 42º encontro na Universidad Pública de Navarra, Pamplona, em 2016, que será liderado por Fermin Mallor. Foram também apresentadas duas candidaturas para a organização da conferência ORAHS 2017, uma pela University of Bath (Christos Vasilakis) e outra pela Shanghai Jiao Tong University (Andrea Matta e Xiaolan Xie).

Maria Eugénia Captivo, Inês Marques, Faculdade de Ciências, Universidade de Lisboa, **Margarida Vaz Pato,** Instituto Superior de Economia e Gestão, Universidade de Lisboa

Para além do programa científico, decorreu o programa social onde não faltou um passeio de barco (naturalmente, no rio Tejo) e um dia de visita à praia da Nazaré e à vila de Óbidos. O jantar oficial da conferência realizou-se na Tenda do Cristo Rei e houve ainda tempo para um muito interessante jantar opcional, na Cozinha Popular da Mouraria, que reuniu mais de metade dos participantes.

O próximo encontro do grupo ORAHS vai realizar-se na HEC Montréal, entre os dias 19 e 24 de julho de 2015, onde todos nos encontraremos de novo, certamente, e de muito bom agrado.

País	Nº Participantes
Alemanha	7
Austrália	1
Áustria	2
Bélgica	6
Brasil	4
Canadá	12
China	2
Colômbia	3
Emirados Árabes Unidos	2
Espanha	7
Estados Unidos da América	4
Finlândia	1
França	6
Holanda	7
Irão	1
Israel	2
Itália	10
Japão	1
Noruega	3
Portugal	33
Reino Unido	23
Sérvia	2
Singapura	1
Turquia	3

Tabela 1: Distribuição do número de participantes por país de origem

IO & BD-2014: WORKSHOPS "IO E BIG DATA"

João Miranda, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Portalegre

O clima de incerteza no ensino superior tem continuado a desenvolver-se, sendo complementado com a rápida evolução tecnológica. Os docentes e investigadores têm observado alterações sucessivas no enquadramento legal, financeiro, económico das suas atividades, considerando-se que o enquadramento global é importante para a definição das estratégias locais, bem como para a evolução da oferta educativa e das principais linhas de investigação que decorrem dos desenvolvimentos internacionais.

A excelência das nossas instituições, a sua posição nos rankings, deriva diretamente dos atributos dos nossos recursos humanos e materiais. A atualização permanente é necessária para manter as dinâmicas existentes no trio investigação-inovação-educação, com ações de caráter geral ou de formação específica, tornando-se também premente a modernização de equipamentos e ferramentas tecnológicas. As tecnologias emergentes em *Big Data, Mobile, Cloud,* redes sociais, têm um papel importante a desenvolver na especialização inteligente duma dada temática, sendo também promotoras da eficiência dos agentes do setor público.

Assim, uma série de workshops sobre a Investigação Operacional (IO) em ambiente de *Big Data* (BD), foram desenvolvidos em colaboração entre a Associação Portuguesa de Investigação Operacional (APDIO), a IBM Portugal e o Instituto Politécnico de Portalegre (IPP), tendo decorrido na Escola Superior de Tecnologia e Gestão (ESTG/IPP), Portalegre, em setembro e outubro de 2014.

Dada a juventude do tema BD em Portugal, preferiu iniciar-se estas atividades com uma jornada onde se reuniram vários oradores em painéis temáticos, sendo os temas selecionados pela sua relevância atual: Saúde, Energia/Indústria, Logística/TIC. Desta forma permitiu-se uma visão geral da situação presente, dos desenvolvimentos futuros decorrentes de aplicações

BD, abordando-se em cada tema as vertentes da IO, das empresas, da inovação/internacionalização. Após este enquadramento geral, desenvolveram-se algumas jornadas laboratoriais visando desenvolvimentos computacionais, aplicações reais, e sendo dirigidas a ferramentas específicas de BD.

Assim, considerou-se no primeiro dia (09-setembro-2014) um encontro de docentes, investigadores, estudantes, e profissionais com apresentações breves em temas relevantes, a que se seguiu um período de discussão. Consideraram-se os seguintes painéis temáticos:

- Saúde com moderação de Adriano Pedro (ESS/IPP), contou com a interação dinâmica de Cristina Semião (Healthcare & Public Sector, IBM Portugal), Maria Eugénia Captivo (FCUL) e Nuno Garcia (UBI);
- Energia/Indústria devido a motivos de força maior, foi ponderada a re-calendarização deste painel;
- Logística/TIC o painel teve os preciosos contributos de Ana Póvoa (IST/UL), José Eduardo Fonseca (*Strategic Outsourcing & Service Delivery*, IBM Portugal) e Luís Vaz (*Business Intelligence*, Grupo Jerónimo Martins), sendo moderado por Valentim Realinho (ESTG/IPP).

Seguidamente, em 10-setembro e 08/09-outubro-2014, decorreram jornadas de formação hands--on com ferramentas IBM (Hadoop, Big Data), visando-se aplicações laboratoriais e casos de estudo. Estas ações de formação específica tiveram a eficiente apresentação e monitorização de Paul Reeves (IBM France).

Embora a frequência nas várias ações decorresse de forma autónoma, o acompanhamento integral das diferentes sessões foi recomendado aos participantes interessados em desenvolvimentos práticos e aplicações BD, tendo-se atingido um total de inscrições na ordem das quatro dezenas e sendo os materiais provisoriamente disponibilizados em http://tiny.cc/IOBD_Sep14.

A Comissão Organizadora das jornadas é constituída por elementos da ESTG/IPP, os quais têm experiência na organização de eventos científicos e de programas internacionais, e.g., http://tiny.cc/sc_sem2014. Para mais detalhes sobre estas jornadas, consultar o resumo alargado, diponível na páginal web da APDIO (http://apdio.pt/home).

A temática em "IO e Big Data" prossegue, estando também calendarizado o IO2015-XVII Congresso da APDIO para 07-09 de setembro-2015, na ESTG/IPP, em Portalegre. O início de inscrições para o IO2015 é esperado em breve, estando informação adicional disponível em http://www.io2015.ipportalegre.pt e sugerindo-se a consulta regular das atualizações.

Além do oportuno suporte da APDIO e da IBM, sem o qual tais jornadas não seriam possíveis, é também de reconhecer o apoio do Centro de Informática da ESTG e dos meios de comunicação da GC/ESTG e do GCRP/IPP. Agradece-se ainda a colaboração do Grupo Jerónimo Martins nas nossas atividades.

GAME CHANGERS: ESTAMOS PREPARADOS?

José Pinto Paixão, Faculdade de Ciências, Universidade de Lisboa

Julgo ser incontestável a emergência de novos fenómenos tecnológicos relacionados com a potenciação da automatização e da robótica, com o acesso generalizado à informação e a passagem da utilização dos computadores da conclusão de tarefas de regra explícita para a análise de vasta e variada informação, com reconhecimento de padrões e desenvolvimento de estratégias de actuação. O desenvolvimento dos campos de aplicação destas inovações pode conduzir ao que se convencionou designar por "second machine age", título de um livro inspirador publicado no início deste ano [2]. Segundo os autores, na "1ª idade", as máquinas funcionaram fundamentalmente como complemento dos humanos substituindo-os sobretudo nas tarefas manuais e repetitivas. Na"2ª idade", registam o reforço da componente de substituição dos próprios humanos em actividades de natureza cognitiva e, mesmo, criativa.

Se assim for, e tudo aponta que seja verdade em vários casos, antecipam-se impactos significativos na sociedade, tanto na economia como nas organizações e nas pessoas.

Acredito que assim será ainda que não me considere a pessoa mais habilitada para descrever os referidos impactos. Nesta breve nota de reflexão, apenas desejo acrescentar um segundo nível de impacto que se relaciona com o conhecimento e, em particular, com a ciência. Especificamente, tentarei evidenciar o papel central que a Investigação Operacional pode desempenhar face aos desafios da "second machine age".

Para isso, focalizo-me em dois dos mais incontroversos novos fenómenos em causa, já bem visíveis nos dias de hoje. Refiro-me à digitalização generalizada (documentos, músicas, vídeo, fotos, notas pessoais,...) e à aquisição intensiva de dados (tráfego, clima, ambiente, saúde, indústria, I&D,...). É impressionante o volume de dados (estruturados, semi-estruturados e não estruturados) que, via *internet* e outros modos de comunicação, estão acessíveis a nível global envolvendo uma grande variedade de indivíduos, grupos e organizações. É o que se identifica genericamente como *big data*, só por si gerador de impactos significativos ao nível do conhecimento científico.

A título de exemplo e procurando fazer a passagem à Investigação Operacional, merece referência um artigo recentemente publicado no boletim trimestral da McKinsey [4], em que se afirma que a aplicação de técnicas analíticas avançadas conjuntamente com lean management podem valer ganhos enormes para os grandes fabricantes mundiais. A Figura 1, retirada do artigo, é bem explícita sobre o impacto estimado desta combinação em vários sectores industriais, sendo de salientar a intervenção da IO em qualquer uma das técnicas analíticas mencionadas nas entradas horizontais do quadro. É óbvio que tal requer a conjugação com outras áreas do conhecimento, como é expresso no artigo que passo a citar explicitamente e apenas com alguns sublinhados pessoais: "This approach (me-Ihoramento contínuo) typically requires forming a small team of econometrics specialists, <u>operations research</u> <u>experts</u>, and statisticians familiar with the appropriate tools. By connecting these analytics experts with their frontline colleagues, companies can begin to identify opportunities for improvement projects that will both increase performance and help operators learn to apply their lean <u>problem-solving skills</u> in new ways".

Efeito Big Data: alguns exemplos

O exemplo de interdisciplinaridade atrás referido replica-se em muitas outras situações que abrangem os mais variados campos de aplicação. Sem pretender ser exaustivo, nem atribuir prioridades, menciono a logística, a mobilidade, a energia e os cuidados de saúde, como exemplo de sectores onde a importância da metodologia, das técnicas analíticas e da IO sobressai quando se perspectiva a extracção de conhecimento com base na análise da informação disponível (big data analytics).

Começo pela **logística** onde a IO tem constituído um importante instrumento de apoio à tomada de decisão. Um estudo de opinião realizado pela Accenture [1], junto de mais de 1 milhar de decisores seniores, revela que 97% dos inquiridos têm expectativas bem fundamentadas sobre os benefícios que a introdução de *big data analytics* pode trazer para as suas cadeias de abastecimento, nomeadamente no que se refere à antecipação do futuro baseada em modelos prescritivos e não apenas descritivos das ocorrências no passado. Contudo, sublinho que o estudo permite concluir que muitas das empresas ainda não possuem uma compreensão clara sobre a forma como o fazer com vista a conseguir um melhor desempenho.

Também na área da mobilidade surgem justificadas expectativas sobre o impacto positivo que a big data analytics pode provocar, designadamente com a criação de novos modelos de negócio e de gestão das infraestruturas. Um exemplo paradigmático é a possibilidade de obter um conhecimento mais correcto e actualizado sobre o relacionamento dos padrões de movimento diários das pessoas envolvendo relações espaciais e temporais entre localizações e actividades muito diferentes. No passado recente, este conhecimento suportava-se essencialmente em resultados de inquéritos, muito exigentes nos recursos alocados, de âmbito restrito e limitados no tempo. Outro exemplo está associado à viabilidade de introdução de sistemas de portagens variáveis em vias e zonas como forma de atenuar o efeito de congestão em determinados períodos do dia e rentabilizar a utilização de infraestruturas.

Na área da **energia**, pode-se afirmar que são já evidentes alguns dos efeitos da informação disponível nos dias de hoje, sobretudo no que se concerne com a



MONTE CARLO SIMULATION

Dynamically identifies bottlenecks and impact of interventions

PRODUCTION AND DISTRIBUTION PLANNING

 $Determines\ optimal\ product\ allocation\ across\ combinations\ of\ substitutable\ production\ lines,\ given\ technical\ limits,\ opportunity\ costs,\ and\ service\ levels$

CAPACITY PLANNING

Evaluates where additional capacity is required when link between resources and output is unclear/complex

VALUE-IN-USE MODELING

Measures impact of different grades of raw materials across integrated value chain

DEMAND AND PRICING OPTIMIZATION

Optimizes SKU³ volumes across regional production centers, given each SKU's net profit contribution to given demand center

PROCESS PARAMETER OPTIMIZATION

Evaluates optimal technical parameters for improved productivity in the chemical-synthesis process $\,$

¹ Extrapolation over top 5 companies in each sector, assuming tools are applicable to only 30% of areas covered within a company and that ~3% additional earnings before interest, taxes, depreciation, and amortization (EBITDA) are unlocked.

² Eletric power and natural gas.

³ Stock-keeping unit.



eficiência energética. No caso particular do consumo de energia doméstica, a crescente introdução de redes de leitores de consumo de energia eléctrica com capacidade de armazenamento e comunicação, permite obter volumes inusitados de dados que, devidamente explorados, podem conduzir a reduções significativas de custos para cada consumidor. Por outro lado, na medida em que uma parte significativa do consumo energético se deve aos transportes, podemos dizer que este sector beneficia da utilização de *big data analytics* nos dois sectores acima mencionados.

Investigação Operacional

Em qualquer das áreas de aplicação acima mencionadas, pode-se afirmar que a IO, ao longo da sua história de existência como domínio de saber, tem tido uma intervenção notável com resultados socioeconómicos evidentes. Aliás, a mesma afirmação é verdadeira para muitas outras áreas da vida real, em que a metodologia e técnicas de IO têm constituído elementos fulcrais para a tomada de decisão aos vários níveis: operacional, táctico e estratégico. Naturalmente, é de esperar que os maiores impactos ocorram no plano estratégico. Na realidade, estão em causa períodos de tempo mais longos que acentuam as características fundamentais do *big data*, emblematicamente designadas pelos seis Vs (volume, velocidade, variedade, variabilidade, valor e veracidade)[3]. Não querendo entrar em grandes especulações, apenas direi que acredito que será a este nível que a IO tem o maior dos desafios decorrentes do efeito. Um desafio que remete a IO às suas "origens", isto é, à capacidade de desenvolvimento de abordagens sistémicas em ambiente interdisciplinar e vocacionadas para a resolução de problemas.



CLINICAL ANALYTICS

- What are the key health indicators across my patient/member population?
- What are the provider quality scores?
- · What is the total cost of care?
- What is our productivity and resource utilization?

ADVANCED ANALYTICS

- What are the main predictors for readmission?
- Which patients/members are most at risk for a bad outcome or ongoing treatment?
- How can I intervene to incentivize patients/ members to make batter choices?
- What treatment programs lead to the best outcome for this patient/member?

Figura 2: Ganhos resultantes da adopção de técnicas de analise prescritiva nos cuidados de saúde

Como último exemplo, quero referir o caso dos cuidados de saúde que ouso antecipar como um dos sectores onde assistiremos a um dos maiores efeitos da análise da informação digitalizada. Na verdade, estima-se que cerca de 80% dos dados médicos são não estruturados, provenientes quer de fontes tradicionais (hospitais, laboratórios, centros de imagem, equipamentos de diagnóstico, etc.) quer de outras fontes não tradicionais, como as redes sociais ou registos de serviços. A Figura 2, retirada de um WhitePaper da IBM [5], identifica os principais ganhos que a adopção de técnicas de análise prescritiva pode implicar. Note-se que esses benefícios vão muito para além da lógica de optimização de recursos, do ponto de vista do operador. De facto, como ressalta na figura, é notório o impacto directo para o paciente, através de um acompanhamento personalizado, seja pela antecipação de necessidades de cuidados médicos, seja pelo desenho e monitorização de tratamentos específicos.

No entanto, para vários sectores, como os que citei explicitamente, o efeito *big data* impõe novas lógicas e objectivos, em qualquer dos níveis de actuação.

Por exemplo, ao nível operacional é expectável que o enfoque das empresas no serviço ao cliente (tempo de resposta, flexibilidade na procura, individualização dos produtos, etc.) obrigue a alterações nos modelos tradicionais, centrados na gestão das operações internas. Contudo, é admissível que tal poderá cingir-se a um quadro disciplinar relativamente bem definido (optimização, estatística, simulação, etc.).

No nível táctico, em geral caracterizado pelas soluções parciais e localizadas, as alterações surgem como consequência do objectivo de integração em redes de serviços. Isto é, obriga a abordagens interdisciplinares em que funções como o planeamento da produção e distribuição perdem rigidez e ganham um dinamismo muito superior ao do passado.

APDIO

Os novos fenómenos tecnológicos (robotização, automação, digitalização, etc.) suscitam interrogações legítimas sobre os correspondentes impactos no plano social e económico. Mas, como procurei mostrar, também geram reflexos no domínio do conhecimento convocando a comunidade científica para uma reflexão sobre os desafios colocados pela "second machine age".

De modo muito sucinto, podemos afirmar que o mundo de hoje enfrenta desafios cada vez mais disruptivos que apelam ao surgimento de "game changers", capazes de desenvolver e implementar soluções de ruptura. Desafios que se colocam em vários planos, nomeadamente, económico, mercado de trabalho, relações sociais, qualidade de vida. Desafios que, naturalmente, se estendem ao campo do conhecimento científico.

A IO pode desempenhar um papel crucial na construção das respostas que a sociedade procura para enfrentar os desafios que colocam em causa muito do que, até ao presente, se entendia como estável. Em concreto, acredito que a comunidade portuguesa de IO tem condições para aproveitar a oportunidade, fazendo uso dos pontos fortes que a distinguem para além da competência reconhecida internacionalmente. Refiro-me, em particular, ao espírito de cooperação interna e à capacidade de intervenção internacional que a APDIO tão bem tem sabido preservar. Estou, pois, convicto que a APDIO poderá assumir um papel mobilizador que nos posicione como "game changers" a nível global. Vamos agarrar a oportunidade?

Referências

- [1] Accenture, Big data analytics in supply chain: hype or here to stay?, Accenture Global Operations Megatrends Study, 1-20, 2014.
- [2] Brynjolfsson, E., McAfee, A., The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies, W. W. Norton & Company Inc., 2014.
- [3] Demirkan, H., Dal, B., Why do so many analytics projects fail?, Analytics-Informs, July/August, 44-52, 2014.
- [4] Dhawan, R., Singh, K., Tuteja, A., When big data goes lean, McKinsey Quarterly, 97–102, 2014.
- [5] IBM, Data-driven health care organizations use data analytics for big gains, IBM Software White Paper, 1-8, 2013.

Filipe Carvalhe

"PENSO QUE O BIG DATA É UMA ENORME OPORTUNIDADE PARA A COMUNIDADE DA INVESTIGAÇÃO OPERACIONAL."





Licenciou-se e obteve o grau de Mestre em Estatística e Investigação Operacional na Faculdade de Ciências da Universidade de Lisboa. Finalizada esta etapa, em 2003 decide fundar a Wide Scope e não seguir o percurso mais natural de procurar uma posição numa empresa já existente. Sendo a generalidade dos portugueses tradicionalmente pouco empreendedora e aversa ao risco, qual a grande motivação para esta sua decisão? Quais as principais dificuldades que encontrou no arranque da Wide Scope?

A Wide Scope foi fundada, por mim e pela Ana Sofia Pereira, com o objectivo de mudar a forma como o mundo opera. Quem começa o seu projecto de empreendedorismo procura ocupar uma posição nova, encontrar o seu lugar. Não tenho a certeza que os portugueses sejam pouco empreendedores, pelo contrário. Talvez tenhamos menos casos de sucesso mas penso que existem muitas e boas iniciativas de uma forma geral.

Ao conhecermos a forma como as empresas operam em muitas das suas actividades mais complexas observámos que poderíamos melhorar drasticamente a sua eficiência, fazendo o mesmo de forma diferente. As técnicas de Investigação Operacional são a ferramenta que permite atingir tal distinção. Por exemplo, como planear as rotas de viaturas de forma a que os consumos de combustível sejam os menores possíveis.

As dificuldades começam com a resistência natural de gestores que sempre fizeram de uma forma manual. O tecido empresarial português é maioritariamente familiar. Ou seja, são empresas estabelecidas há muito tempo, por vezes em gerações anteriores, geralmente bem sucedidas ao longo desse tempo.

Tendo as operações sempre sido feitas da mesma forma e a empresa tido sucesso é assumido que essa forma de trabalhar não deve ser alterada. Esta visão sustentada pelo sucesso passado ostraciza a inovação futura dentro da empresa. E este foi um dos principais obstáculos para a Wide Scope que, apesar de apresentar tecnologias e novas formas de operar, precisou de vencer muitos obstáculos para se estabelecer. Ter a melhor solução não é suficiente.

A Wide Scope é uma empresa que desenvolve software para optimização de processos na cadeia logística e de produção, sendo o mérito do seu trabalho reconhecido quer em Portugal, quer no exterior. A comprová-lo estão algumas das distinções recebidas, nomeadamente o prémio IAPMEI Excelência para as PME em 2012, tendo sido incluída na lista de Cool Vendors in Supply Chain Management Applications 2014 pela prestigiada consultora Gartner. Considerando a actual situação económica, quais os factores que identifica como indispensáveis para todo este sucesso?

Em todas as adversidades devemos procurar oportunidades. Se um dos maiores obstáculos que encontrámos foi a abertura para fazer melhor e racionalizar custos, toda a situação económica desfavorável abriu a receptividade a esses temas. Preferia que a crise não tivesse acontecido, mas uma boa gestão tem de saber navegar com todos os ventos.

Não concorremos a prémios mas têm-nos sido atribuídos reconhecimentos ímpares. A Wide Scope é a única empresa portuguesa alguma vez referida pela Gartner na área de *Supply-Chain*, habitualmente reservada a empresas norte-americanas. Também a Deloitte assinalou a nossa performance ao longo de 5 anos consecutivos como uma das 500 empresas de tecnologias de informação de maior crescimento na Europa, Médio-Oriente e África.

É um objectivo difícil para uma empresa de Investigação Operacional, com uma oferta de nicho, e a trabalhar em Portugal durante uma crise financeira. Os factores de diferenciação estão na inovação dos produtos da Wide Scope e também na estratégia de gestão, que assenta num posicionamento único.

Hoje em dia as grandes empresas têm acesso a grandes quantidades de informação, em diferentes formatos e em actualização permanente, sendo frequente ouvirmos falar de *Big Data*. Surge assim o desafio de extrair conhecimento relevante para o negócio de cada empresa a partir deste grande volume de informação. Sendo a Wide Scope uma empresa preocupada em acompanhar a vanguarda da ciência e tecnologia, qual a experiência que já dispõe nesta área?

A Wide Scope está a tirar partido de *Big Data* em várias aplicações. Por exemplo, a nossa solução de optimização de rotas inclui uma componente de trânsito preditivo que permite prever como estará o trânsito amanhã numa determinada estrada. Com *Big Data* podemos ter em conta a previsão do tempo (chove?), se há greve de transportes anunciada, se é dia de início de aulas e também todo o histórico de tráfego nessa estrada com dados recolhidos a partir dos localizadores GPS dos veículos ou mesmo dos sinais de telemóveis. É um volume enorme de dados de diversas fontes cujo tratamento em tempo útil só é possível usando técnicas de *Big Data*.

Considerando uma empresa que deseje aderir ao paradigma do Big Data, coordenando os seus actuais sistemas de informação com outro tipo de dados não estruturados por forma a aumentar o seu nível de conhecimento sobre o negócio, quais as preocupações chave que deve ter presente na respectiva implementação?

A principal preocupação é a qualidade dos dados. Se estamos a falar de dados em elevado volume convém que tenham qualidade, sob pena de estarmos a exponenciar ruído. Este é o maior desafio para as empresas que habitualmente têm dados transacionais de alguma qualidade, mas nada mais. Ou seja, sabem quanto venderam ontem mas não têm ideia sobre o que não

venderam, quais as páginas por onde um cliente que desistiu navegou no seu *website*, etc.

Penso que *Big Data* não é um fim em si, mas apenas um meio para atingir um fim. Ou seja, a preocupação não deve estar em encontrar um lugar para a tecnologia mas sim em perceber como nos podemos servir da tecnologia para suprir necessidades identificadas.

A gestão deste novo tipo de dados obriga a uma colaboração estreita entre diferentes domínios da ciência, com especial relevância para a Informática e a Matemática. Que desafios se colocam? Que papel pensa que a Investigação Operacional pode desempenhar nesta nova realidade? Penso que o Big Data é uma enorme oportunidade para a comunidade da Investigação Operacional. O perfil mais crítico de um projecto de Big Data é o Data

Scientist, ou seja, o especialista que sabe o que fazer com os dados e que conclusões inferir. As áreas de machine learning e optimização ficam assim sobrevalorizadas. Os skills necessários não se encontram na área das tecnologias de informação mas sim no domínio da matemática.

Ao longo do tempo começaremos a assistir a ferramentas que já incluirão muitos algoritmos, procurando minimizar a dependência de um *Data Scientist*, mas isso abrirá outras oportunidades do ponto de vista de utilização destas ferramentas. Dificilmente serão operadas por pessoas sem sensibilidade matemática.

OUTRAS NOTÍCIAS



A sessão pública de apresentação do livro Investigação Operacional em Ação - Casos de Aplicação, coordenada pelos Professores Doutores Rui Oliveira e José Soeiro Ferreira, teve lugar no passado dia 15 de Dezembro de 2014, na Casa da Cultura de Coimbra, com apresentação do Engenheiro Carlos Moreira da Silva.

Usaram da palavra o director da Imprensa da Universidade de Coimbra, Professor Delfim Leão, o apresentador convidado Engenheiro Carlos Moreira da Silva, os coordenadores da obra Professor Rui Oliveira e Professor José Soeiro Ferreira, e o presidente da Comissão Directiva da APDIO Professor Domingos Cardoso.

Esta obra integra uma colecção de casos de aplicação da Investigação Operacional em Portugal, relatando actuações reais com forte ligação empresarial, que constituem um complemento à formação académica, tradicionalmente mais enfocada na resolução de *canned problems*.

Pelo quinto ano consecutivo, foi atribuído o Prémio APDIO - FCT UNL ao melhor aluno de IO da Faculdade de Ciências e Tecnologia da UNL, que foi constituído com parte dos lucros do IO2009 - 14º congresso nacional da APDIO, que decorreu na FCT-UNL. O prémio relativo ao ano lectivo 2013/14 foi atribuído *ex-aequo* a **Ricardo Jorge da Conceição Silva**, aluno da Licenciatura em Matemática e a **Cidália Margarida Machado Tomás**, aluna do Mestrado em Matemática e Aplicações daquela faculdade. Os diplomas e cheques correspondentes ao prémio foram entregues na celebração do Dia da FCT-UNL, que teve lugar no passado dia 12 de Novembro de 2014, na referida faculdade.

TESE DE DOUTORAMENTO CONCLUÍDA RECENTEMENTE

Autor: Mahdi Dodangeh

Título: Worst case complexity of direct search under convexity

Instituição: Faculdade de Ciências e Tecnologia – Universidade de Coimbra

Designação do Doutoramento: Programa Inter-Universitário de Doutoramento em Matemática Data de conclusão: Novembro de 2014

Orientador: Luís Nunes Vicente

EVENTOS APOIADOS PELA APDIO

Mini EURO Conference on Improving Healthcare: new challenges, new approaches

30 de Março a 1 de Abril de 2015 Coimbra, Portugal

http://www.minieuro2015.com

EVENTOS ORGANIZADOS PELA APDIO

Courses Mixed-Integer Programming/Disjunctive Programming - Mixed-Integer Models for Planning, Scheduling

27 e 28 de Janeiro de 2015 Lisboa, Portugal **http://apdio.pt**

IO2015 - XVII Congresso da APDIO

7 a 9 de Setembro de 2015 Portalegre, Portugal

I or talegre, i ortagai

http://www.io2015.ipportalegre.pt

Qualquer contribuição para o Boletim deve ser enviada para boletim.apdio@campus.fct.unl.pt

A (R)EVOLUÇÃO DO BIG DATA

Rodrigo Rodrigues, Departamento de Informática, Universidade Nova de Lisboa

O termo *Big Data* refere-se a um conjunto de técnicas e de sistemas informáticos que permitem o processamento em paralelo de grandes volumes de dados. Estes sistemas têm hoje um papel muito relevante na análise da quantidade cada vez maior de dados que são recolhidos quer por empresas, quer no contexto de atividades de investigação, e cuja análise é um fator cada vez mais importante para estas organizações.

As origens desta área de Big Data podem atribuir-se ao sistema MapReduce, concebido pela Google. Em particular, este sistema foi da autoria de dois dos seus principais engenheiros, Jeff Dean e Sanjay Ghemawat, e o seu funcionamento foi descrito através de uma publicação científica em 2004, numa das conferências mais importantes da área dos sistemas computacionais, o Symposium on Operating Systems Design and Implementation [1]. Nesse artigo é explicado que vários sistemas tinham sido construídos até à data para efetuar cálculos e computações sobre grandes volumes de dados, nomeadamente sobre o conjunto de todas as páginas da web, que eram periodicamente lidas e armazenadas em servidores da Google. O MapReduce surge assim como uma forma de estancar a profusão de sistemas que vinham a ser desenvolvidos especificamente para cada uma destas computações. Assim, o MapReduce teve o mérito de fornecer uma plataforma única para permitir que vários computadores trabalhem em paralelo para processar os grandes volumes de dados armazenados nas mesmas máquinas. Adicionalmente, o MapReduce tem várias características que tornam a sua utilização pelos programadores das computações paralelas muito acessível. Em particular, o MapReduce tem uma interface muito simples, e lida também com as falhas dos computadores que fazem parte do sistema de forma eficiente e transparente, isto é, sem que o programador da computação paralela tenha de escrever código para lidar com estas mesmas falhas. Estas características tornaram o conceito muito atrativo, e, volvido apenas um ano sobre esta publicação, outros dois engenheiros, um dos quais a trabalhar na Yahoo!, desenvolveram uma versão em código fonte aberto (open source) da plataforma MapReduce a que chamaram Hadoop [2], e que é hoje muito utilizada quer a nível académico quer a nível das empresas e outras organizações.

A receção do paradigma MapReduce para a computação paralela por parte de algumas das comunidades de investigação da área da informática esteve longe de ser calorosa. Tal deveu-se ao facto de, à altura, estas comunidades trabalharem com outros sistemas que permitiam atingir objetivos semelhantes. Em particular, sistemas como o MPI (Message Passing Interface) ou os sistemas paralelos de gestão de bases de dados ofereciam algumas vantagens sobre o paradigma MapReduce e a plataforma Hadoop, nomeadamente ao nível do desempenho e da possibilidade de

expressar computações complexas de forma concisa. No entanto, a simplicidade de programar computações paralelas no paradigma MapReduce e a facilidade de instalar e usar a plataforma Hadoop, bem como a forma eficiente e transparente como esta lida com as falhas, tornaram o MapReduce e o Hadoop o *standard de facto* para efetuar computações paralelas sobre grandes volumes de dados.

O funcionamento do MapReduce é relativamente simples. O programador de uma computação paralela começa por armazenar os dados de entrada dessa computação num sistema de armazenamento, que tipicamente é um sistema especificamente concebido para armazenar grandes volumes de dados de forma distribuída (o Hadoop fornece para tal um sistema que faz parte da respetiva distribuição chamado HDFS). Depois, o programador apenas tem de completar duas funções específicas à sua computação: a função Map que lê um subconjunto do input e emite um conjunto de pares <chave, valor>, e a função Reduce que agrega todos os valores emitidos para cada chave, e escreve como saída do programa um conjunto de valores finais que voltam a ser guardados no sistema de armazenamento. Um exemplo canónico de uma computação neste paradigma é o cálculo de um histograma com a ocorrência de palavras num documento, em que a função Map lê uma parte do documento e para cada palavra encontrada p emite o par <p,1>, ao passo que a função Reduce conta o número de pares recebidos para cada palavra e tem como saída o valor final dessa contagem. Para implementar esta funcionalidade, o sistema MapReduce (e a plataforma Hadoop) lançam um conjunto de tarefas que podem ser escalonadas em qualquer máquina do conjunto de servidores dedicado a esta computação. Estas podem ser tarefas Map, que invocam a função Map sobre um subconjunto dos dados de entrada, ou tarefas Reduce, que, após a conclusão de todas as tarefas Map, correm a função Reduce para uma dada chave, passando como parâmetros todos os valores emitidos pelas tarefas Map, e associados a essa chave. Um processo escalonador central orquestra quais as tarefas que devem ser lançadas a cada instante e em que máquinas devem correr, e lida também com as falhas. A forma de lidar com falhas no MapReduce é muito simples, e tira partido do facto de o trabalho de cada tarefa poder ser repetido em qualquer máquina sem prejudicar o resultado da computação. Desta forma, ao detetar que um servidor não consegue ser contactado, o escalonador pode simplesmente reiniciar as tarefas pelas quais esse servidor era responsável noutra máquina, sem qualquer outro tipo de diligência para recuperar de uma possível execução incompleta das tarefas em causa.

Nos anos que se seguiram ao aparecimento do MapReduce e do Hadoop, assistiu-se ao aparecimento de muitas outras plataformas que aperfeiçoaram alguns aspetos do MapReduce ou se especializaram em determinados tipos de processamento paralelo, como por exemplo o processamento de dados que vão sendo continuamente produzidos, e exigem uma constante atualização do resultado da computação à medida que novos dados são produzidos. Este cada vez maior ecossistema de plataformas de "Big Data" torna difícil a escolha da plataforma mais apropriada para cada tipo de computação, mas abre também portas a uma maior aplicação deste tipo de processamento em problemas de diversas áreas. Entre as áreas de aplicação das plataformas de "Biq Data" encontra-se a Investigação Operacional, dado que muitos problemas de simulação, de otimização, ou outros problemas relacionados com esta área têm soluções, muitas vezes aproximadas, que são extremamente exigentes sob o ponto de vista computacional. Existem já algumas bibliotecas disponíveis com vários algoritmos programados para a plataforma Hadoop. Um exemplo de uma biblioteca de grande utilidade é a Apache Mahout [3], que disponibiliza, por exemplo, algoritmos para fatorização de matrizes ou de clustering.

Em suma, e no momento em que a plataforma MapReduce celebra a primeira década de existência, a lição principal a reter é que o MapReduce surge como o percursor da área de "Big Data" por ser a plataforma certa - dada a sua simplicidade e facilidade de utilização - na altura certa - dado o momento histórico em que há muitos dados para processar e cada vez mais organizações têm acesso a um vasto número de servidores onde podem correr computações paralelas. No que diz respeito a este último aspeto, é de salientar que hoje em dia qualquer indivíduo ou organização pode ter acesso a um grande número de servidores sem sequer ter de investir na compra das máquinas. Tal deve-se à possibilidade de recorrer a serviços de cloud computing, onde estes recursos computacionais podem ser alugados apenas pelo tempo necessário.

Para o futuro, será de esperar que este tipo de plataformas ganhem cada vez mais utilizadores e sejam aplicadas em cada vez mais áreas. Assim, e da mesma forma que até agora os conhecimentos de programação eram essenciais para se poder trabalhar na área da Investigação Operacional, será de esperar que no futuro a programação para plataformas de "Big Data" possa também tornar-se um pré-requisito para trabalhar esta e outras áreas. Desta forma, a inclusão de tópicos de "Big Data" nos respetivos currículos académicos poderá fazer sentido num futuro não muito distante.

Referências

- [1] Dean, J., Ghemawat, S., MapReduce: Simplified data processing on large clusters, OSDI'04: Sixth Symposium on Operating Systems Design and Implementation, December, 2004.
- [2] http://hadoop.apache.org
- [3] http://mahout.apache.org



PORTUGUESES EM IO PELO MUNDO

Miguel Anjos,

Full Professor e Canada Research Chair (em Discrete Nonlinear Optimization in Engineering), Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Canadá

Começo por agradecer à Ana Luísa Custódio e à Isabel Correia a oportunidade de contribuir para o Boletim da APDIO. Espero que este comentário sobre a minha carreira e a IO seja de algum interesse para os restantes colegas da APDIO.

Nasci no Porto e completei em Portugal o ensino primário, antes de emigrar para o Canadá. A maior parte dos meus estudos foi assim realizada em Montréal, até um nível equivalente ao de uma licenciatura no atual sistema de Bolonha. Seguidamente fiz um mestrado nos Estados Unidos, com uma bolsa concedida pelo governo canadiano. Depois de completar o mestrado, regressei a Montréal e aceitei emprego na direção de uma empresa industrial. Após trabalhar três anos no ramo industrial, tomei consciência da falta que sentia do meio universitário e da investigação científica que realizara durante três verões passados a trabalhar como assistente-investigador. Decidi então fazer um doutoramento para poder sequir a carreira universitária.

Foi assim que em 2001 me doutorei em otimização na Universidade de Waterloo, sob a supervisão do Professor Henry Wolkowicz. (Como o Joaquim Martins, também eu tenho dificuldade em escrever otimização sem o "p"!) Especializei-me em otimização não-linear com variáveis inteiras, mais conhecida pela abreviação inglesa MINLP, e especificamente na aplicação da otimização semi-definida positiva para melhor aproximar certos problemas de MINLP. Após concluir o doutoramento, trabalhei alguns anos na Alemanha e no Reino Unido antes de regressar ao Canadá, primeiro a Waterloo, Ontário, para depois voltar a Montréal, fechando assim um ciclo de quase 15 anos desde o princípio do doutoramento.

Para além do facto evidente de ter seguido ou aproveitado as diversas possibilidades de emprego que me foram sendo oferecidas, vejo retrospetivamente neste percurso vários aspetos positivos. Descobri, por experiência própria, que no contexto académico existem grandes vantagens em estudar e trabalhar em meios diferentes, colaborar com vários grupos e contactar com visões diversas da mesma área de investigação. Um dos exemplos mais frutuosos de colaboração que tenho é com o grupo do Professor Michael Jünger da Universidade de Colónia, na Alemanha. Integrei este grupo pela primeira vez em 2001-2002, como investigador pós-doutoral, e regressei em 2009-2010 como Fellow Humboldt. No decorrer deste período publiquei vários artigos com membros do grupo e a colaboração continua, não só com o grupo em Colónia mas também com alguns dos seus membros que mudaram para outras universidades e dirigem agora os seus próprios grupos de investigação.

Foi também através de colaborações que as minhas atividades de investigação se expandiram para além da otimização semi-definida positiva, passando a incluir aplicações da otimização não-linear. Este ramo tem aumentado em popularidade nos últimos anos, por exemplo no sector da energia elétrica, com as Smart Grids, onde a planificação e operação da rede elétrica de forma ótima são temas de atualidade. Uma das questões importantes é a integração das fontes de eletricidade eólica e solar cuja geração intermitente ocasiona múltiplos desafios técnicos e económicos, alguns dos quais podem ser aprofundados através de modelos não-lineares do comportamento da rede sob condições variáveis. Estas fontes intermitentes são um dos principais meios para "descarbonizar" a rede elétrica e assim reduzir a emissão de gases responsáveis pelo efeito estufa, um dos principais temas das atividades do Institut de l'Énergie Trottier, da Polytechnique Montréal, que dirijo desde o seu lançamento em meados de 2013.

Uma das maiores influências na minha carreira foi a organização canadiana Mitacs. Fundada há 15 anos como Rede Canadiana de Centros de Excelência para a Investigação nas Ciências Matemáticas, a Mitacs transformou-se desde então numa organização que promove colaborações de investigação entre empresas e universidades, através do financiamento de estágios para estudantes. O meu envolvimento com a Mitacs começou

quando recebi apoio financeiro como estudante de doutoramento através de um projeto intitulado High Performance Optimization, liderado pelo Professor Tamás Terlaky. Alguns anos depois de completar o doutoramento fui convidado para participar neste projeto como professor-investigador, tendo acabado por ser co-líder do mesmo. Além da minha colaboração como investigador, desde 2011 que faço parte do Conselho de Investigação da Mitacs, cuja responsabilidade é prestar liderança científica assim como elaborar avaliações críticas dos vários programas. Um dos aspetos mais importantes dos programas Mitacs é a obrigação do estagiário passar um mínimo de 50% da duração do estágio na empresa, sob a supervisão de um co-director que não é investigador universitário. A minha experiência pessoal, assim como a de muitos estagiários Mitacs, é que esta obrigação leva a uma autêntica integração do estagiário na empresa. Isto não só contribui para o sucesso do projeto mas também leva muitas vezes à contratação do estagiário. O modo de funcionamento da Mitacs é muito bem sucedido e o envolvimento de empresas neste tipo de programa tem aumentado a um ritmo anual impressionante.

O futuro da nossa profissão está diretamente ligado com o seu sucesso em aplicações nas áreas de engenharia, gestão ou quaisquer outros ramos fora da nossa comunidade, onde o valor da IO pode ser demonstrado e apreciado. Ao mesmo tempo, para ter sucesso na carreira de investigador operacional, creio que é vantajoso ter uma especialização sólida numa das metodologias da IO. A minha especialização em otimização não-linear permite sempre posicionar-me dentro de projetos interdisciplinares com uma identidade clara, que indica imediatamente a contribuição que terei no quadro do projeto. Os investigadores operacionais que adquiram uma formação metodológica sólida, que estejam disponíveis para deslocar-se em função das oportunidades e para participar em projetos interdisciplinares, terão um futuro promissor.

BIG DATA E DATA SCIENCE

Luís Cavique, Universidade Aberta

1. Introdução

A informatização dos serviços, desde as sofisticadas transações em bolsa à simples compra de um café, associada às redes sociais e aos dispositivos móveis (tablets, smart-phones) produzem uma enorme quantidade de dados. Para além da quantidade de dados, a taxa de atualização desses mesmos dados é também enorme. Atualmente, em cada 10 minutos são gerados mais dados do que todos os dados gerados desde a pré-história até ao ano de 2003.

Por outro lado, a capacidade de processamento também tem tido aumentos significativos. Nos últimos 40 anos a capacidade de integração dos circuitos integrados permitiu duplicar em cada dois anos a capacidade de processamento [14], aumentar a capacidade de armazenamento e reduzir o respetivo preço. A lei de Moore tem-se verificado nos últimos 40 anos, permitindo um aumento no processamento na ordem de 2²⁰.

Para ter uma noção do aumento da capacidade, dada a dificuldade dos humanos para compreender o significado de um crescimento exponencial, vamos usar o exemplo do tempo de viagem de Lisboa ao Porto. Suponhamos que uma viagem de Lisboa ao Porto, há 40 anos atrás, decorria em média em 6 horas. Se a evolução dos transportes fosse tão grande como nos computadores, o tempo da viagem de Lisboa ao Porto, hoje em dia, teria a duração de 2 centésimos de segundo (6x60x 60/2²º).

O grande volume de dados compensado pelo aumento da capacidade de processamento tem originado novos conceitos, como o *Big Data* e a criação de novas profissões como os *data scientists*, apelidada pela Harvard Business Review como a profissão mais *sexy* do século XXI.

2. Big Data

Com o advento da web 2.0 (a web das pessoas) associada aos dispositivos móveis e à *internet of things*, as clássicas aplicações empresariais foram largamente ultrapassadas em volume de dados.

Num estudo realizado em 2012, o valor estimado de informação no planeta foi de 2,8 ZB (zetabytes, 10²¹ bytes). A mudança de escala no volume de dados e na sua taxa de atualização deu origem ao que genericamente se chama de *Big Data*.

Ao nome *Big Data* está associada a sigla 3V: volume, velocidade de atualização e variedade dos formatos. Alguns autores incluem um quarto V de valor ou veracidade dos dados.

Dos 2,8 ZB existentes, 85% são dados não estruturados, ou seja, *medias* como o vídeo, fotografia e som. Dos restantes 15%, de dados formatados e de texto, só 3% são analisados. Concluímos assim que só a pequena percentagem de 0,45% dos dados do planeta são objeto de análise. Tal como na astrofísica, onde a matéria negra contém eventualmente várias explicações para a origem do universo, os 99,55% dos dados não analisados são apelidados de *dark data*.

Com o surgimento de novos formatos de dados estruturados surgiu dentro do *Big Data* o conceito do NoSQL [3]. O NoSQL, ou Notonly SQL, permite o armazenamento, tratamento e consulta de dados de forma muito eficiente. As soluções NoSQL estão divididas em alguns grupos:

- Armazenamento de chave/valor, como Voldemort da LinkedIn.
- Armazenamento de super-colunas, como HBase ou Cassandra do Facebook.
- Armazenamento de documentos, como XMLdatabase ou MongoDB.
- Armazenamento de grafos, como HyperGraphDB ou ArangoDB.
- Armazenamento orientado a objetos, como Db4obiect.

Tendo como ponto de partida o modelo relacional e a linguagem declarativa SQL (*Structured Query Language*), utilizada na maior parte das bases de dados das empresas, o NoSQL apresenta-se como a alternativa para lidar com grandes volumes de dados.

As estruturas das soluções NoSQL foram simplificadas relativamente ao modelo relacional e garantem a consulta da informação de forma muito eficiente, com complexidades algorítmicas de ordem O(1).

Em NoSQL a complexidade máxima deve ser da ordem O(N), sendo N a dimensão do ficheiro. Ao contrário do SQL a operação de junção de tabelas não existe, dada a sua elevada complexidade de ordem O(N²) para o pior caso.

Para a agregação de dados é utilizado o conceito de *MapReduce*, implementado em duas fases. A função do operador *Map* seleciona os dados em subgrupos. A operação *Reduce* agrega a informação de cada subgrupo. A complexidade algorítmica no pior caso será de duas vezes O(N).

O conceito de *Big Data* traz um conjunto de novos desafios para lidar com grandes volumes de dados, tanto para as empresas como para a comunidade científica. O desenvolvimento de novos algoritmos é crítico já que as complexidades algorítmicas são de preferência de ordem O(1) e nunca devem exceder a ordem O(N).

Por consequência, o *Big Data* cria novas oportunidades na tomada de decisão baseada em dados, *data driven decisions*. Tal como refere Peter Norvig, diretor

da Google Research, "nós não temos melhores algoritmos; nós temos mais dados" [10].

3. Data Science

Data Science, é o atual termo para a ciência que analisa dados, combinando a estatística com machine learning/data mining e tecnologias de base de dados, para responder ao desafio que o Big Data apresenta.

O termo criado na década de 2010, *Data Science*, corresponde aquilo que nos anos de 1970 se apelidava de *Decision Support Systems*, DSS, nos anos 80 aos *Executive Information Systems*, EIS, nos anos 90 aos *Online Analytical Processing*, OLAP, e nos anos de 2000 ao *Business Intelligence*, BI [10].

3.1 Base de Dados versus Data Mining

As questões colocadas que têm resposta numa Base de Dados são semelhantes às questões colocadas ao analista de *Data Mining*.

Em Base de Dados pretende-se por exemplo:

- Identificar os clientes que compraram mais de 1000 euros.
- · Identificar os dois produtos mais vendidos.
- Identificar os 10 clientes com mais reclamações.
- Enquanto que em Data Mining procura-se:
- Identificar os grupos de clientes com hábitos de compra idênticos (clustering).
- Encontrar o produto X que é adquirido com o produto Y (regras associativas).
- Encontrar os atributos que levam os clientes a reclamar (classificação).

Embora as questões sejam semelhantes, nas Bases de Dados é apresentado um padrão (e.g. consulta SQL) e são devolvidos dados, por outro lado, em *Data Mining* são fornecidos os dados e pretende-se extrair padrões.

3.2 Macro e Micro padrões

Data Science é um processo computacional para descobrir "padrões". Padrão é uma forma com uma configuração específica e facilmente reconhecível, que se caracteriza por uma regularidade, repetição de partes e acumulação de elementos. Por exemplo, uma duna criada pela ação do vento é composta por várias camadas de areia e tem uma configuração reconhecível.

Os micro-padrões correspondem a pequenas percentagens de dados; por exemplo nas regras associativas, uma medida de suporte que apresenta valores suporte ≥ 5%, sendo escolhidas as regras com maior *confidence* (ou probabilidade condicionada). Por outro lado, os macro-padrões envolvem uma grande percentagem, ou a totalidade, dos dados; por exemplo na modelação com regressão são utilizados todos os dados disponíveis. Os micro-padrões caracterizam-se por

elevada confiança (confidence) e os macro-padrões por elevado suporte.

Existem outros exemplos de micro-padrões: nos problemas de sequence/episode mining com suporte maior ou igual a 1%; no problema de classificação, ao utilizar decision trees, cada ramo da árvore corresponde a uma pequena percentagem dos dados; ainda no problema de classificação ao utilizar o k-nearest neighbor a comparação que é efetuada é com o reduzido número de k elementos. Quanto aos macro-padrões, em problemas como regressão, teste de hipóteses, clustering ou redução de atributos, todos os dados são tidos em consideração.

A origem desta dicotomia na análise de dados remonta aquando do aparecimento do *Data Mining*, hoje uma área madura, mas que tinha inicialmente uma conotação negativa com os nomes de *data snooping* (bisbilhotando) e *data fishing*, onde o objetivo era explorar e/ou espiar subconjuntos de dados.

Leo Breiman em 2001 [5] já tinha referido as duas culturas na modelação de dados. A cultura dos micro-padrões corresponde à procura de pequenas percentagens de dados com eventual utilidade ou interesse. Esta abordagem tem tido, até à data, um grande apoio dos grandes decisores dos EUA em projetos de mais de 1.000.000 dólares. A cultura dos macro-padrões utiliza a totalidade dos dados, tem origem na matemática e na estatística e conta com projetos vinte vezes menores que os anteriores.

No atual paradigma de *Big Data*, em que as complexidades algorítmicas não devem exceder O(N), grande parte dos algoritmos de *machine learning/data mining* são desadequados. A reutilização das métricas da estatística combinada com a tecnologia de base de dados faz anunciar a reconciliação das duas culturas na modelação de dados na recente *Data Science*.

3.3 Data Scientist

Data Scientist é apelidada como a profissão mais sexy do século XXI [11]. Por data scientist entende-se "alguém melhor em estatística que um engenheiro informático e alguém melhor em programação do que um matemático". A maior parte dos programadores não se querem envolver em conceitos matemáticos e da mesma forma os estatísticos não aceitam programar em SQL, R ou Python.

O data scientist será assim alguém que saiba diferenciar um teste de hipóteses t-student de um qui-quadrado, ao mesmo tempo que sabe ver a diferença entre um algoritmo polinomial de ordem O(N) e de O(N²).

As duas culturas na modelação de dados criaram especialistas que atualmente são obrigados a compatibilizar esforços. A Investigação Operacional encontra-se numa posição privilegiada já que sempre combinou os conceitos da matemática com a sua aplicação nas ciências informáticas.

4. Redução da Dimensionalidade

Em ambientes *Big Data* o volume é grande, dinâmico e não estruturado. Por outro lado, não existem algoritmos disponíveis para responder a este desafio. O volume de dados não analisados (*dark data*) é ao mesmo tempo uma oportunidade e uma inquietação, visto que os dados gerados excedem largamente a capacidade de armazenamento instalada.

Se não se pode alterar de imediato a complexidade dos algoritmos, a resposta pode estar na redução da dimensionalidade dos dados. Em Investigação Operacional existe uma larga experiência neste campo. A análise das componentes principais e a análise fatorial são técnicas conhecidas na estatística e em *machine learning* para redução do número de variáveis.

A redução da dimensão pode ainda ser realizada pela transformação do problema e pela sumarização dos casos (ou linhas). Os dados são condensados com vista a encontrar padrões de grandes subconjuntos de dados, utilizando portanto a abordagem dos macro-padrões. De seguida apresentamos vários exemplos de transformações em redes e grafos.

4.1 Análise Topológica de Dados

A Análise Topológica de Dados [6] representa os dados utilizando redes. A rede agrupa dados semelhantes em nós e cria arcos se existe partilha de dados entre dois nós diferentes. Visto que cada nó representa vários pontos, a rede permite comprimir os dados com uma alta dimensionalidade para uma representação de mais baixa dimensionalidade.

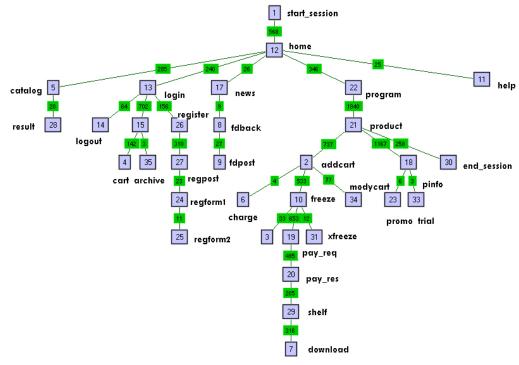


Figura 1: Resultado do algoritmo Ramex num problema de web mining.

A topologia é o estudo da forma, em que se distinguem três propriedades que relaxam o conceito de forma: a invariância das coordenadas, a invariância da deformação e a compressão das representações. A invariância da deformação é particularmente interessante, já que se adapta perfeitamente à capacidade dos humanos de compreenderem as formas. Por exemplo, a letra "A" é compreendida pelos humanos qualquer que seja a fonte tipográfica utilizada. Estas características fazem com que a Análise Topológica de Dados se esteja a tornar numa nova área de estudo em data mining e na visualização de dados.

4.2 Process Mining

Uma percentagem apreciável do Big Data corresponde aos históricos de eventos (eventlogs) que são gerados em cada instante, durante 7 x 24 horas, por milhares de milhões de dispositivos fixos e móveis.

Process Mining [1] é uma técnica que permite extrair informação de históricos de eventos. Segundo o autor, o Process Mining cria pontes entre o Data Mining e o Business Process Modeling. A técnica considera a acumulação de eventos tendo como objetivo melhorar a representação dos dados. A abordagem utiliza redes de Petri, onde cada processo é representado por um nó, e as sequências de eventos podem ser condensadas na rede. Esta abordagem cria novos padrões para os problemas antigos de sequence mining, tendo em vista criar equilíbrios entre a simplicidade e a exatidão dos resultados.

4.3 Similis

As regras associativas ficaram célebres, ao encontrar um padrão nos supermercados onde jovens casais com filhos às sextas-feiras e sábados, ao comprar fraldas também compravam cerveja. A regra fraldas => cervejas, tendo uma métrica de suporte de algumas centésimas, estava associada a uma confiança (probabilidade condicionada) relevante.

O algoritmo Apriori [2] foi o primeiro algoritmo para o Market Basket Analysis. O Apriori gera para um pequeno número de produtos um enorme conjunto de regras associativas, i.e. micro-padrões, que devem ser criteriosamente escolhidas pelo utilizador final. O trabalho de Cavique [7] com o algoritmo Similis, resolve o Market Basket Analysis para um elevado número de produtos e evita a escolha entre milhares de micro--padrões, devolvendo padrões baseados na acumulação, i.e. macro-padrões que representam a totalidade dos dados. O algoritmo Similis está dividido em duas partes. Na primeira transforma o problema num grafo ponderado e na segunda encontra subgrafos completos que correspondem aos cabazes de compras mais comprados.

4.4 Ramex

A descoberta de padrões sequenciais é um tema muito importante em data mining, dado o grande número de aplicações que incluem a análise de compras, web mining, sequência de ADN, entre outros.

O algoritmo inicial, o AprioriAll [15], para além de ter uma elevada complexidade temporal, encontra milhares de micro-padrões de difícil seleção e que requerem um trabalho exaustivo na atribuição de utilidade ou interesse.

As cadeias de Markov representam um conjunto de estados associados com um conjunto de transições entre estados. No caso da análise do cabaz de compras, cada estado corresponde a um item e no caso da navegação da web, cada estado é uma página. Os modelos de Markov foram usados para representar e analisar os utilizadores que navegam na web em [4].

No problema de sequence mining tratado por Cavique [8], é apresentado o algoritmo Ramex que gera árvores que envolvem todos os elementos numa perspetiva de macro-padrões. Ramex tem origem no latim e significa ramos de uma árvore. Na abordagem pretendemos manter a visão global dos itens e evitar tempos computacionais não-polinomiais. Usando heurísticas baseadas no algoritmo da árvore geradora, podem ser encontradas em redes cíclicas as estruturas de árvores com maior peso, que correspondem aos padrões sequenciais mais frequentes.

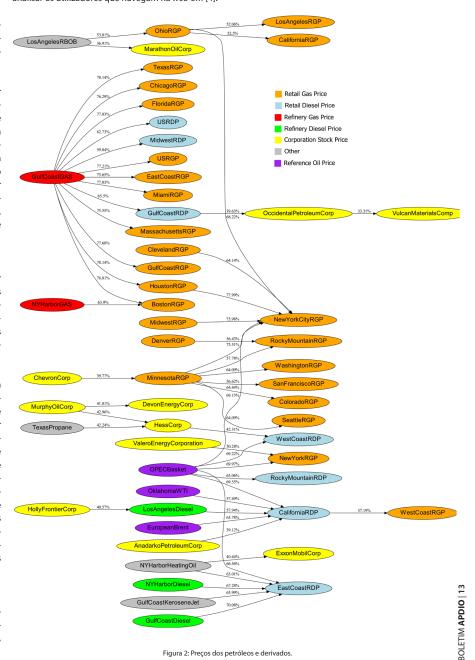


Figura 2: Preços dos petróleos e derivados.

O algoritmo Ramex divide-se em duas fases. Numa primeira fase o ficheiro de entrada é transformado num grafo orientado. Na segunda fase é aplicado o algoritmo *Maximum Weight Rooted Branching* definido por Fulkerson [12]. Neste algoritmo é dado um nó inicial a partir do qual se desenvolve uma árvore.

O algoritmo foi testado num problema de web mining tendo sido encontradas as sequências da Figura 1. Cada ramo da árvore corresponde a uma sequência de eventos com ramos idênticos à estrutura do web site.

4.5 Ramex com poli-árvores

Uma poli-árvore é um grafo orientado acíclico com um arco entre cada par de nós no máximo. O grau interno dos vértices de uma árvore é zero (a raiz) ou um. Por sua vez, o grau interno dos vértices de uma poli-árvore pode ser maior que um. Podemos ainda acrescentar que numa poli-árvore para cada par de nós só pode existir uma única sequência de nós.

No trabalho [9] é apresentada a versão do Ramex utilizando poli-árvores para a deteção de padrões sequenciais. Para testar a nova abordagem foram usados ficheiros de grandes dimensões. As experiências foram implementadas utilizando os ficheiros gerados pelo IBM Quest Synthetic. O algoritmo utiliza uma matriz inicial semelhante às das Cadeias de Markov, mas usa uma heurística polinomial baseada no algoritmo de Prim para determinar os padrões. Nesta abordagem encontramos as seguintes vantagens:

i) Incremental: Visto que os dados dos eventos são transformados em pesos no grafo, a atualização de novos eventos pode ser realizada de forma incremental.

ii) Inexistência de parâmetros: A maior parte dos algoritmos para deteção de sequências utilizam o suporte mínimo com o parâmetro para controlar a explosão combinatória. Para o algoritmo proposto não há necessidade de qualquer parâmetro.

iii) Escalabilidade: Em comparação com os demais algoritmos, a nossa abordagem não faz uma procura exaustiva. Contudo, utiliza os dados condensados numa rede. O procedimento que devolve o resultado da árvore tem uma complexidade polinomial e apresenta uma ótima escalabilidade.

iv) Visualização: Usualmente os pacotes de software mais conhecidos geram um grande número de regras, perdendo-se portanto a visão global. Na nossa abordagem todos os itens são tomados em consideração e a visualização das poli-árvores mais pesadas corresponde ao raio-X das sequências de eventos.

A utilização do algoritmo Ramex aplicado aos mercados financeiros deu origem ao Ramex-Forum [13]. A Figura 2 apresenta os resultados do algoritmo para as influências dos preços dos petróleos e derivados, extraída do trabalho de Tiple [16].

5. Conclusões

Neste artigo foram apresentados os conceitos básicos de *Big Data* e a nova área a que deu origem, a *Data Science*. Em *Data Science* foi discutida e exemplificada a noção de redução da dimensionalidade dos dados.

Como conclusões para a IO em ação, podemos referir duas grandes oportunidades que o *Big Data* oferece:

i) A Investigação Operacional encontra-se numa

situação privilegiada, ao combinar, desde sempre, a matemática e a informática, para lidar com o *Data Science* e para liderar a formação numa das profissões mais atraentes do século XXI. A necessidade de voltar a recorrer aos algoritmos de baixa complexidade da estatística, coloca as técnicas de IO na vanguarda.

ii) O ambiente Big Data exige aos programadores e investigadores um conjunto de novos algoritmos, tornando-se urgente a redução da complexidade temporal de quase todos os algoritmos, desde o simples cálculo da variância, em estatística, até ao mais complexo problema de sequence mining. Para responder a este desafio a redução da dimensionalidade é uma abordagem já demonstrada. Os exemplos apresentados utilizam duas fases distintas. A primeira fase acumula os dados em bruto numa estrutura de dados condensados: rede na Análise Topológica de Dados [6], rede de Petri [1], cadeia Markov [4] ou grafo [8]. Na segunda fase é possível procurar os macro-padrões na estrutura de dados condensados. Os algoritmos para as referidas estruturas de dados são igualmente conhecidos na Investigação Operacional, tornando este tipo de redução da dimensão dos problemas muito aliciante.

Referências

- [1] Aalst, W. van der, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer-Verlag Berlin Heidelberg, 2011.
- [2] Agrawal, R., Srikant, R., Fast algorithms for mining association rules, Proceedings of the 20th International Conference on Very Large Data Bases, 487-499, 1994.
- [3] Alexandre, J., Cavique, L., NoSQL no suporte à análise de grande volume de dados, Revista de Ciências da Computação, 8, 37-48, 2013.
- [4] Borges, J., Levene, M., Evaluating variable-length Markov chain models for analysis of user web navigation sessions, IEEE Transactions on Knowledge and Data Engineering, 19, 441-452, 2007.
- [5] Breiman, L., Statistical modeling: the two cultures, Statistical Science, 16, 199-231, 2001.
- $[6] \ Carlsson, G., Topology \ and \ data, \textit{Bulletin of the American Mathematical Society}, 46, 255-308, 2009.$
- [7] Cavique, L., A scalable algorithm for the market basket analysis, Journal of Retailing and Consumer Services, Special Issue on Data Mining Applications in Retailing and Consumer Services, 14, 400-407, 2007.
- [8] Cavique, L., A network algorithm to discover sequential patterns, *Progress in Artificial Intelligence, EPIA 2007, Lecture Notes in Computer Science*, 4874, J. Neves, M. Santos e J. Machado (eds.), Springer-Verlag Berlin Heidelberg, 406-414, 2007.
- [9] Cavique, L., Coelho, J. S., Descoberta de padrões sequenciais utilizando árvores orientadas, Revista de Ciências da Computação, 3, 12-22, 2008.
- $[10] \ Davenport, T.\ H., \textit{Big Data at Work: Dispelling the Myths, Uncovering the Opportunities}, Harvard\ Business\ School\ Publishing\ Corporation, 2014.$
- [11] Davenport, T. H., Patil, D. J., Data scientist: the sexiest job of the 21st century, Harvard Business Review, 90, 70–76, 2012.
- [12] Fulkerson, D. R., Packing rooted directed cuts in a weighted directed graph, Mathematical Programming, 6, 1-13, 1974.
- [13] Marques, N. C., Cavique, L., Sequential pattern mining of price interactions, EPIA 2013, 16th Portuguese Conference, Advances in Artificial Intelligence, Local Proceedings, Angra do Heroismo, Açores, Portugal, 314-325, 2013.
- $[14] \, Moore, G.\,E., Cramming \, more \, components \, onto \, integrated \, circuits, \textit{Electronics}, \, 114-117, \, 1965.$
- [15] Srikant, R., Agrawal, R., Mining sequential patterns: generalizations and performance improvements, Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, EDBT, Lecture Notes in Computer Science, 1057, 3–17, 1996.
- [16] Tiple, P. S., Tool for Discovering Sequential Patterns in Financial Markets, Dissertação para obtenção do Grau de Mestre em Engenharia Informática, Faculdade de Ciências e Tecnologia da Universidade Nova Lisboa, 2014.

DISTRIBUIÇÃO DE GPL A GRANEL: CASO DE ESTUDO DA GALP ENERGIA

Tatiana Gonçalves de Sousa, Maria Cândida Mourão, Instituto Superior de Economia e Gestão. Universidade de Lisboa

A expressão "Time is money, and distance matters" representa o que se faz sentir na área da distribuição de todo o tipo de produto, em qualquer empresa. Neste âmbito, têm surgido inúmeros estudos com o intuito de otimizar as rotas de distribuição, visando a redução dos custos associados à logística. Segundo o *Eurostat*, relativamente a dados de 2012, Portugal detinha uma das mais extensas redes de autoestradas da União Europeia [1]. Na Figura 1 pode-se observar a proporção do total de bens transportados, em Portugal, por segmento de transportes, no ano de 2013. Como se pode verificar, 41% do total de bens foi transportado por via rodoviária.



Figura 1: Proporção do total de bens transportados, em 2013, por seg-

Neste contexto, surgiu este estudo, que a empresa portuguesa Galp Energia (http://www.galpenergia.com) se disponibilizou a "apadrinhar", com o objetivo de apurar se haveria espaço para melhorias nas rotas de distribuição de GPL.

O principal objetivo do presente estudo prende-se com o desenvolvimento de um programa em VBA, com o intuito de apoiar a tomada de decisão por parte dos responsáveis da distribuição. Ambiciona-se que este programa permita otimizar as rotas de distribuição, tendo--se para tal desenvolvido um modelo matemático.

Pretende-se desta forma diminuir as distâncias percorridas na distribuição do GPL, assegurando a satisfação dos pedidos efetuados pelos diversos clientes. Em termos práticos, o objetivo fundamental que se pretende alcançar é o desenvolvimento de um programa que possibilite minimizar a distância total percorrida por um veículo cisterna. Esta é calculada pelo percurso efetuado desde a partida do veículo cisterna do parque de armazenamento, passando pelos locais de entrega do produto aos clientes, até ao retorno ao parque de partida.

As restrições e pressupostos a considerar são:

- A procura tem de ser sempre satisfeita, independentemente da quantidade de GPL.
- É inevitável respeitar as capacidades máximas de transporte de GPL, decretadas na lei, para cada um dos tipos de veículos cisterna.
- Cada cliente é visitado uma só vez em cada dia, por um único veículo cisterna.
- Cada rota começa e termina num e num só parque de armazenamento. Não irá pois considerar-se a possibilidade de o ponto de origem poder diferir do ponto de retorno.
- Os veículos cisterna apenas realizam tarefas de distribuição.

O primeiro passo passou pela recolha e análise da informação disponível sobre as rotas de GPL. Os dados recolhidos estão relacionados com o tipo de produto, veículos, procura de cada cliente, parques de armazenamento, zonas genericamente identificadas e distâncias entre os clientes. Para efeitos de análise, optou-se por consultar os registos compreendidos entre janeiro e março de 2014, uma vez que são dados com pouco desfasamento temporal, e incluem os meses de maior consumo de GPL.

De modo a tornar o processo mais célere, desenvolveu-se um programa em VBA que automatiza o processamento dos dados fornecidos pela Galp Energia, e que permite a ligação com o software solver premium (http://www.solver.com/premium-solver-platform), utilizado na identificação das rotas ótimas.

Seguidamente, encontram-se retratadas as categorias nas quais ocorreu a recolha dos dados e, posteriormente, o tratamento dos mesmos.

Tipos de produto

Embora a operação de distribuição da Galp Energia abranja diferentes tipos de produtos petrolíferos, aqui refere-se apenas o GPL a granel.

Veículos cisterna

A informação dos veículos que constituem a frota da empresa, embora não seja expressa no modelo de otimização de rotas, será importante para os resultados e conclusões finais.

Disponibilização de informação

A informação sobre as rotas efetuadas pelos diversos veículos é agregada num aparelho designado por *Portable Document Format* (PDT). Através deste mecanismo, estabelece-se a comunicação entre a empresa e o motorista e, desta forma, a informação

relevante chega aos escritórios da empresa. A comunicação de informação é tratada e enviada através de um ficheiro em Excel – designado por SGL, onde se encontra retratada toda a atividade de distribuição de um determinado período, inclusive as quantidades de GPL transportadas para cada cliente.

Parques de armazenamento

A Galp Energia detém três parques de armazenamento de GPL: o parque de Perafita, a Companhia Logística de Combustíveis (CLC), e a refinaria de Sines.

Distância entre clientes

Para otimizar o número de quilómetros a percorrer pelos veículos cisterna, foi feito um levantamento das distâncias entre os clientes, e entre estes e os parques de armazenamento de GPL. Com o auxílio de um programa escrito em VBA, foi possível a recolha das distâncias referidas a partir da base de dados dos clientes, sendo os ficheiros SGL percorridos automaticamente.

Através de uma aplicação da *Google* – a *Google Maps API* – foi possível implementar em VBA um método que indica quer as distâncias, quer os tempos de deslocamento, e os escreve numa matriz de origens e destinos.

Este passo foi um dos mais complexos e exaustivos, uma vez que a base de dados dos clientes da Galp Energia não se encontrava no formato adequado aos processos utilizados, o que representou uma grande limitação para o programa desenvolvido. Por este facto, foi necessário despender um tempo considerável a modificar as moradas, com o intuito de chegar a valores o mais próximo possível dos valores reais.

Após ter sido reunida e tratada a informação indispensável, verificou-se a necessidade de definir e formular um *Vehicle Routing Problem* (VRP) [2] como estratégia de resolução do caso de estudo.

Ao tratamento dos dados, seguiu-se a resolução do problema real através da aplicação do modelo matemático desenvolvido, que, como referido, se baseia num problema de planeamento de rotas (VRP). Neste estudo consideraram-se 85 dias de trabalho, durante os meses de janeiro a março. Os meses foram escolhidos por serem de inverno, logo, os de maior consumo de GPL. Os resultados foram segmentados por zonas geográficas, de acordo com os três parques de armazenamento existentes. Para este período foram disponibilizados 43 veículos de diferentes categorias. No mês de janeiro foram registados 28 dias de

distribuição e as rotas foram efetuadas por 38 dos 43 veículos: 16 cisterna do tipo 2 eixos, 19 do tipo 3 eixos e 3 semirreboques.

No mês de fevereiro recorreu-se a 36 veículos: 13 de 2 eixos, 20 de 3 eixos e 3 semirreboques. As rotas efetuadas pelos veículos neste mês foram percorridas durante 26 dias.

Por último, em março, foram registados 31 dias de distribuição. Nestes, foram utilizados 41 veículos: 16 de 2 eixos, 22 de 3 eixos e 3 semirreboques.

Na Figura 2 comparam-se os valores das distâncias das rotas obtidas face às utilizadas pela empresa (denominadas por "reais"), de acordo com a zona geográfica estudada.

Observou-se que as rotas percorridas representam um acréscimo de, aproximadamente, 11% (desvio nas 3 zonas geográficas) face às rotas calculadas pelo modelo de otimização. Na zona norte e na zona centro observou-se um acréscimo de sensivelmente 12%. A zona onde o percurso escolhido mais se aproximou da rota ótima foi a zona sul do país, onde se observou um desvio de apenas 9%. Esta diferença deve-se, essencialmente, às particularidades das rotas nessa área geográfica, pois a região do Algarve dispõe de menos rotas alternativas, quando comparada com as restantes zonas analisadas.

Constatou-se, também, que 57% dos fretes efetuados nos três meses em análise tiveram um desvio entre 0 e 10% face às rotas ótimas. Ainda assim, subsiste uma percentagem de 44% de fretes que ultrapassa a bartermos da satisfação das necessidades dos clientes. Neste contexto surge o desenvolvimento do presente estudo com o intuito de otimizar as rotas de distribuição de GPL da empresa portuguesa Galp Energia, permitindo fortalecer a eficiência económica, sempre com enfoque no cliente.

Dada a complexidade do problema em estudo, foi necessário efetuar algumas simplificações. De modo a reduzir a complexidade do modelo não foram consideradas restrições de trânsito, como por exemplo a que se refere ao horário na Ponte 25 de Abril e a existência de túneis ou vias em que não é permitida a passagem a viaturas de transporte de mercadorias perigosas (facto esse que afetaria os veículos cisterna que transportam GPL). Não se considerou, igualmente, a possibilidade de o ponto de origem diferir do ponto de retorno e as situações em que é necessário recolher sobras de gás.

Outra limitação prende-se com a aplicação da Google que permite o cálculo das distâncias e do tempo de viagem requerido. Esta aplicação, implementada no código em VBA, tem as suas próprias limitações de uso. De facto, cada consulta enviada à API da Google é limitada pelo número de elementos permitidos, apurado, neste caso, pelo produto entre o número de origens e o número de destinos. Atualmente, apenas se permitem 100 elementos por consulta, 100 elementos a cada 10 segundos e 2 500 elementos a cada período de 24 horas.

Em conclusão, a aplicação do modelo matemático permitiu definir as rotas ótimas, o que resultaria numa diminuição da distância total a percorrer. Tudo isto conduz a uma importante diminuição dos custos logísticos. Note-se que, apesar da simplificação do problema, o presente estudo mostra o quão importante se torna investir num bom programa de otimização de rotas, uma vez que mesmo simplificado, a resolução do problema proporcionaria melhorias em termos de custos para a Galp Energia.

COMPARAÇÃO DA DISTÂNCIA TOTAL PERCORRIDA DURANTE OS TRÊS MESES, POR ZONA GEOGRÁFICA, FACE À DISTÂNCIA ÓTIMA

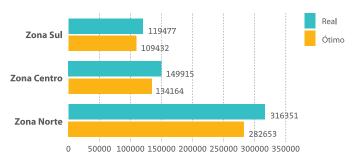


Figura 2: Comparação da distância total percorrida durante os três meses, por zona geográfica.

Energia é feita apenas com base na experiência, sem apoio a qualquer modelo matemático.

Como seria expectável, a distância total percorrida foi superior à que resulta da situação modelada e correspondente otimização de rotas, denotando-se um desvio relativamente às rotas ideais em todos os 85 dias. Assim, a aplicação do modelo ao período em estudo permitiria uma redução de 59 494km no total dos dias, o que se traduz numa poupança de, aproximadamente, 700 km por dia. Esta constatação mostra a importância da aplicação dos problemas de planeamento de rotas a casos reais para a otimização dos recursos utilizados no transporte.

Atualmente, a definição das rotas por parte da Galp reira dos 10%. Pôde, ainda, observar-se que 10% dos fretes efetuados pela empresa percorreram a rota ótima, o que é uma percentagem considerável.

> No que aos custos diz respeito, constata-se que a utilização do modelo desenvolvido permitiria poupanças que rondam os 10,1%. (Por uma questão de sigilo, não será possível mostrar os custos em €/km e, consequentemente, os respetivos custos totais das rotas.)

Conclusões

Os problemas de planeamento de rotas gozam, cada vez mais, de uma importância significativa ao nível empresarial, uma vez que assumem um grande impacto quer em termos de custos logísticos, quer em

Referências

[1] Eurostat. Disponível em: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=ttr00002&plugin=1 (consultado em 07.12.2012).

[2] Toth, P., Vigo, D., The Vehicle Routing Problem, SIAM Monographs on Discrete Mathematics and Applications, 2002.