

Clodis Boscaroli

Análise de Agrupamentos baseada na
Topologia dos Dados e em
Mapas Auto-organizáveis

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do Título de Doutor em
Engenharia Elétrica.

São Paulo
2008

Clodis Boscaroli

Análise de Agrupamentos baseada na
Topologia dos Dados e em
Mapas Auto-organizáveis

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do Título de Doutor em
Engenharia Elétrica.

Área de concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Emilio Del Moral Hernandez

São Paulo
2008

FICHA CATALOGRÁFICA

Boscarioli, Clodis

Análise de Agrupamentos baseada na Topologia dos Dados e em Mapas Auto-organizáveis. São Paulo, 2008. 118 p.

Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1.Descoberta de conhecimento 2.Mineração de dados 3.Análise de agrupamentos 4.Redes neurais I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos. II.t.

À Deus e aos espíritos superiores que me acompanham e orientam nesta vida, e
à minha família, hóspedes eternos no meu coração!

Agradecimentos

Aos meus pais e primeiros mestres, Clélia e Laércio, que me apoiaram em todos os caminhos que escolhi percorrer e que, mesmo sem muitas vezes compreenderem meus propósitos, sempre aceitaram minhas escolhas.

Aos meus irmãos Cássia e Cleiton e ao meu cunhado Antonio, por demonstrarem, de maneira admirável, como é possível manter unida uma família, estando eu sempre geograficamente distante.

Aos meus sobrinhos Camila e Flávio e ao meu afilhado Luka, crianças lindas com quem tanto gostaria de conviver mais, acompanhando o crescimento.

Além da família que amo demais, são tantas as pessoas a agradecer nessa fase de doutoramento, que chega a ser difícil. Sem meus amigos, os de todos os tempos e os conquistados a partir de 2003 - ano em que cheguei em São Paulo, na USP, sem conhecer nada e ninguém, acho que essa jornada não teria sido possível.

Aos ex-professores Marcos Sfair Sunye e Laura Sanchés García (UFPR) e Maria Salete Marcon Gomes Vaz (UEPG), por terem fornecido boas cartas de recomendação, que certamente, foram fundamentais no processo de seleção.

Ao meu orientador, Prof. Emilio Del Moral Hernandez pela confiança depositada, pelo apoio e exemplo de profissional e ser humano. Sempre me lembrarei de você chegando com Virgínia no HU-USP, sábado à noite, após saber que eu seria submetido à cirurgia de fratura de tornozelo. Obrigado pelo companheirismo e preocupação. É um prazer fazer parte do ICONE (Grupo de Inteligência Computacional, Modelagem e Neurocomputação), sob sua orientação.

À Sarajane Marques Peres, amiga de todas as horas e tantas afinidades. Muito importantes foram as nossas longas e sempre proveitosas conversas via Internet durante e sobre nossos doutorados. Obrigado pela sempre pronta disposição em colaborar e por ter-se proposto a ler partes do que veio a tornar-se esse documento.

À minha amiga Lucia Fabíola da Silva. Veterana nos tempos da graduação. Colega. Amiga de trabalho. Amiga. Ex-aluna em pós-graduação *lato sensu*. Amiga. Parceira de viagem. Amiga. Várias foram as formas que interagimos desde que nos conhecemos, mas a nossa amizade sincera e sempre presente só aumentaram.

Aos amigos e compadres Lucio Mauro e Karina Machado e aos amigos Anne Gigliola Silva, Ezequiel Gueiber e Benifrancis Matias. Além de amigos fiéis, vocês são referências constantes pelas quais sempre quero retornar a Ponta Grossa para reencontros.

À Patrícia Junqueira Grandino, grande amiga que tanto admiro também pela trajetória profissional. Amiga de ir ao cinema, sair para dançar e principalmente, de poder contar. Sem palavras de agradecimento pela acolhida em sua família desde o momento em que saí do hospital até voltar a andar. Além de sua amizade, a convivência com Fernanda e Pedro e D. Amélia foi-me um presente.

Agradeço também à Nilza, que de bom grado e carinhosamente incluiu minha recuperação em suas atividades diárias de secretária do lar, ajudando-me “nos tempos do andador” com os banhos de contraste, administração dos remédios etc. Certamente esse cuidado acelerou minha recuperação.

Ao Alexandre César Tavares Vidal, meu irmão de coração e um dos amigos mais cultos e de humor impagável que tenho. Obrigado pelos “tempos de república”, pelo carinho fraterno e pelo seu exemplo de perseverança.

À Priscila Braga Calíope, amiga de laboratório que gosto muito. Obrigado por ter-me apresentado a Emilio e ao ICONE, e pela convivência sempre agradável no laboratório.

Ao “pai dos gêmeos”, meu amigo Marcel Jaques Simonette, pelos nossos “cafés com bobagem”, almoços de papos sérios sobre nossas preocupações, anseios e impressões sobre cursar pós-graduação, que em muito me ajudaram em momentos de desânimo ou cansaço.

Ao grande amigo de laboratório e de “tomar uma cerveja no final do expediente”, Leandro Augusto da Silva, em muito agradeço: pela amizade, pela convivência no Grupo de Pesquisa, por ter revisado parte dessa tese, pela ajuda

com MatLab e pelas tantas horas que estudamos juntos e/ou, discutimos nossas pesquisas de doutorado.

À minha amiga Claudia de Oliveira Melo, por estar sempre presente, mesmo que nos últimos tempos à distância. Nossa empatia foi imediata. Altos papos, bons momentos e apoio mútuo. Além de amigos, temos em comum o fato cabuloso de ambos termos quebrado o pé durante a pós-graduação. Valeu as dicas variadas sobre cuidados na fase de recuperação. Obrigado também pela revisão de alguns capítulos dessa tese.

Ao Marcelo Facio Palin, um grande amigo e um exemplo de bondade, por quem tenho muito carinho e respeito. Agradeço a amizade, as nossas conversas sobre docência e sobre o futuro, e pelas dicas sobre Latex.

Ao casal Diana Maia e Gustavo Esteves, amigos pelos quais nutro um carinho muito grande. Eles foram também os responsáveis pela minha introdução à capoeira.

À Déborah Schneider, Neidinha, Tutty, Sara A. de Medeira, Roberta R. Parciannelo, Ivone Rojo e Luiz Antonio Rodrigues, bons amigos de Cascavel. Retornar tornou-se mais fácil pelo fato de vocês estarem aí. Déborah e D. Ivone, obrigado também por terem me acolhido antes de novamente estabelecer residência em Cascavel.

Aos meus colegas de trabalho da Universidade Estadual do Oeste do Paraná por muito me incentivarem e apoiarem. Em especial, a Jorge Bidarra e Claudia Brandelero Rizzi pela “cobrança” pelo término desse trabalho, sempre evidenciando sua importância ao nosso Grupo de Pesquisa, o GIA (Grupo de Inteligência Aplicada), que caminha para sua consolidação. E à própria Instituição, por ter-me liberado por três anos das atividades docentes para dedicar-me ao Doutorado.

Aos professores com os quais tive contato durante esse período em disciplinas e/ou seminários, o meu muito obrigado. Aprendi muito com vocês e alguns certamente me servirão de exemplo na trajetória acadêmica. Destaco o Prof. Dr. João Eduardo Ferreira, o jef, um entusiasta nos Sistemas de Banco de Dados, com quem muito aprendi nas disciplinas cursadas, como bolsista em Projeto PAE ou nos seminários de seu Grupo de Pesquisa em Bancos de Dados no IME-USP.

Aos funcionários da EP, IME e EACH, pelo carinho e profissionalismo com que sempre me trataram. Destaco entre eles, Claudia Ferreira de Souza Leite, Solange Paixão, Célia Maria de Lima, Marcelo Vantini, Adriano de J. Alves Ferreira, Lucileide Tadei e Maximiano Gonçalves Oliveira. Faço um agradecimento especial à Valquiria Alaminos, que com o tempo tornou-se uma amiga muito querida.

Durante esse período de doutoramento também conheci muitas outras pessoas legais, inteligentes e interessantes, que também cursavam pós-graduação, que fizeram-se presentes nesse período. Entre elas, Arlindo F. Conceição, Daniela Finco, Elier B. Cristo, Ana Beatriz V. Graciano, Jihan Zoghbi, Mehran Misaghi, Giulian D. Luz, Giuliano Mega, Ney Barros Luz, Raphael Y. de Camargo, Silvia Calbo, Silvia Cristina Pinto, Rosianni Cruz, Renato da S. Ramalho, Humberto Sandamann, Celina M. Takemura, Eyder Martinez Montoya, Rodrigo Fresneda, Ricardo L. Andrade, José Ribamar Braga Jr., Karina Valdívia, Kelly Rosa Braghetto, Eduardo Takeo Ueda, Ana K. Simões e Rosimeiri Toyota. A convivência e o compartilhar diferentes experiências foram-me muito gratificantes.

Aos meus amigos das comunidades “Trilheiros & Bêbados”, “Equipe Pé na Lama” e seus agregados, pelas trilhas e contato consciente e saudável com a natureza que ajudaram-me a manter a serenidade. Em especial a Claudio Sekiguti, que literalmente salvou-me a vida, quando em uma queda fiquei pendurado em uma cachoeira. Também aos amigos que saíram das atividades dos grupos para uma grande amizade: Roberval Mela Jr., Kariny Ribeiro, Luiza M. Rocha, Adriana Luchini, Ana Paula Pazini, Suzete Pazini, Andressa V. Robes, Adriana Martins e Gabriel Cordeiro Lima. A preocupação e o cuidado de vocês quando da cirurgia e pós-operatório são inesquecíveis.

Gostaria também de agradecer a todas as pessoas que de alguma forma passaram pela minha vida nesse período. Como não acredito em acasos, certamente esses “encontros” colaboraram para tornar-me a pessoa que hoje sou.

Sinceramente, estou muito feliz ao terminar essa seção de agradecimentos, por ter me dado conta do tanto que eu tenho para agradecer a Deus!

Resumo

Cada vez mais, na conjuntura das grandes tomadas de decisões, a análise de dados massivamente armazenados se torna uma necessidade das mais variadas áreas de conhecimento. A análise de dados envolve a realização de diferentes tarefas, que podem ser realizadas por diferentes técnicas e estratégias como análise de agrupamento de dados. Esta pesquisa enfatiza a realização da tarefa de análise de agrupamento de dados (*Data Clustering*) usando SOM (*Self-Organizing Maps*) como principal artefato. SOM é uma rede neural artificial baseada em aprendizado competitivo e não-supervisionado, o que significa que o treinamento é inteiramente guiado pelos dados e que os neurônios do mapa competem entre si. Essa rede neural possui a habilidade de formar mapeamentos que quantizam os dados, preservando a sua topologia. Este trabalho introduz uma nova metodologia de análise de agrupamentos a partir de SOM, que considera o mapa topológico gerado por ele e a topologia dos dados no processo de agrupamento. Uma análise experimental e comparativa é apresentada, evidenciando a potencialidade da proposta, destacando, por fim, as principais contribuições do trabalho.

Abstract

More than ever, in environment of large decision-making, the analysis of data stored massively becomes a real need in almost all knowledge areas. The data analyzing process covers the performing of different tasks that can be executed for different techniques and strategies like the data clustering analysis. This research is focused on the analysis task of data groups, called Data Clustering using Self Organizing Maps (SOM) as the principal artifact. SOM is an artificial neural network based on competitive and unsupervised learning, what means that its training is entirely driven by the data, such the neurons of the map compete themselves for doing it. This neural network has the ability to build the mapping task that quantifies the source data, but preserving the topology. This work introduces a new clustering analysis methodology based on SOM, considering the topological map produced by it and also the topology of the data obtained in the clustering process. The experimental and comparative analysis are also presented to demonstrate the potential of the proposal, highlighting at the end the main contributions of the work.

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

