

# Proposta de Modelo Adaptativo para Geração de Contextos na Recomendação de Locais

C. V. Crivelaro<sup>1</sup>, F. J. Barth<sup>2</sup>, R. L. A. Rocha<sup>3</sup>

**Resumo**— Sistemas de Recomendação dominaram o mercado, especialmente na Internet para a recomendação de itens para o usuário. No entanto, muitos Sistemas de Recomendação de Locais tratam apenas de um possível contexto do usuário, mesmo este podendo estar em vários contextos ao mesmo tempo, como trabalho e viagens de férias. Esta proposta tem como objetivo propor um modelo adaptativo para a criação de múltiplos contextos para o mesmo usuário e, assim, fazer recomendações mais aderentes.

**Palavras-Chave**— Sistemas de Recomendação, Adaptatividade, Geolocalização

## I. INTRODUÇÃO

Os Sistemas de Recomendação ganharam grande apelo em aplicações comerciais, por recomendar novos itens aos usuários como locais para ir, produtos como livros [1], músicas e até amigos nas redes sociais. Em muitos casos, os Sistemas de Recomendação aumentaram os acessos aos itens dos sites ou mesmo ao número de vendas, como no site de venda de livros Amazon [1].

As recomendações são feitas de acordo com o histórico de navegação do usuário e pela comunidade em que o usuário está inserido. Assim, ele se sente familiarizado com o item oferecido. Neste caso de estudo, serão utilizados os locais como itens para recomendação do site brasileiro Apontador.

No site Apontador, o usuário pode encontrar informações sobre locais em todo o Brasil e fazer avaliações para compartilhar a sua experiência do local com a comunidade. Um dos desafios é apresentar aos usuários os locais mais aderentes ao seu interesse de visitação e que sejam mais próximos do seu cotidiano, como um restaurante diferente para a hora do almoço, um restaurante para o final de semana com os amigos, ou até mesmo locais para visitação durante as férias.

Ao fazer a recomendação de locais ao usuário, infere-se que o usuário vai gostar do local e atribui-se um valor para ordenar os *Top N* locais que lhe serão oferecidos. No entanto, o usuário pode modificar o seu contexto temporariamente como a verificação de locais para uma viagem ou ter dois contextos diferentes, como a passagem de alguns dias durante o meio da semana em uma cidade e o final de semana em outra cidade. Esse fato diminui a qualidade das recomendações, pois, se a distância entre os locais sugeridos e os avaliados for levada em consideração, pode haver recomendações fora do contexto ou fora do foco das sugestões em que o usuário não está mais presente.

Com este problema, esta proposta de trabalho propõe uma solução adaptativa para que o usuário possa ter vários contextos em paralelo à medida que ele visita ou avalia os locais no site e fornece mais informações sobre o seu contexto. Com mais de um contexto do usuário, é possível focar a recomendação no momento em que o usuário estiver dentro de um contexto específico, evitando a mistura de recomendações de locais diferentes.

Esta proposta está estruturada da seguinte forma: na Seção 2 é apresentada uma breve descrição sobre Sistemas de Recomendação; na Seção 3 é fornecida uma definição de Sistema de Recomendação com contexto de locais; na Seção 4 é apresentado um breve resumo sobre Tecnologia Adaptativa; na Seção 5 é apresentada a proposta deste trabalho; e na Seção 6 são apresentadas as medidas para avaliar se essa proposta terá um ganho sobre outras propostas já conhecidas.

## II. SISTEMAS DE RECOMENDAÇÃO

Os Sistemas de Recomendação [2][3] surgiram no contexto de Sistemas de Recuperação de Informações, que consistiam na busca dos itens mais interessantes ao usuário baseado apenas em sua navegação e em seus termos de busca. Esses Sistemas de Recomendação são conhecidos como Sistema de Recomendação Baseados em Conteúdo.

Nessa abordagem, são comparados os itens que o usuário interagiu com os outros itens da base. Quanto mais próximos forem os itens sugeridos dos itens interagidos, melhores pontuações terão para a recomendação, sendo uma abordagem interessante para recomendação de locais parecidos com o que o usuário já visitou. Para fazer essa recomendação, basta tratar os itens como documentos e fazer a similaridade entre eles por meio do modelo de espaço vetorial com pesos TF-IDF (frequência de termos – inverso da frequência do documento). [4]. Entretanto, essa abordagem tem vários problemas devido à recomendação ser feita apenas com os dados do passado do usuário, e não ligar outros locais que não são próximos semanticamente.

No contexto de locais, as pessoas não estão interessadas apenas em lugares iguais aos anteriores que elas visitaram, mas também em lugares em que pessoas que têm mesmo interesse vão. Para isso, há os Sistemas de Recomendação por Filtragem Colaborativa.

Com o surgimento do *Tapestry* [5], surgiu o termo Filtragem Colaborativa, cuja ideia é, se os usuários tiveram os mesmos interesses no passado, terão os mesmos interesses no futuro. Isso implica não só em ligar os itens pelo seu conteúdo, mas ligar os itens ou os usuários pelo interesse. Com esse conceito, podemos ter recomendações de acordo com a comunidade de pessoas que vão aos mesmos lugares e oferecer os lugares que estão em maior destaque no momento

<sup>1</sup> Celso V. Crivelaro – Apontador e Escola Politécnica da USP – [celso.crivelaro@apontador.com](mailto:celso.crivelaro@apontador.com)

<sup>2</sup> Fabrício J. Barth – Apontador – [fabricao.barth@apontador.com](mailto:fabricao.barth@apontador.com)

<sup>3</sup> Ricardo L. A. Rocha – Escola Politécnica da USP – [luis.rocha@poli.usp.br](mailto:luis.rocha@poli.usp.br)

ao usuário que demonstra afinidade com outros usuários. Formalizando o problema, o Sistema de Recomendação faz a sugestão de lugares baseado na matriz Usuários ( $\mathcal{U}$ ) e Itens ( $\mathcal{J}$ ) para gerar uma lista de Recomendações ( $\mathcal{R}$ ), sendo essas recomendações os itens que o usuário ainda não avaliou:

$$\mathcal{U} \times \mathcal{J} \rightarrow \mathcal{R} \quad (1)$$

Um exemplo de uma matriz  $\mathcal{U} \times \mathcal{J}$  pode ser mostrado abaixo:

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$	3	1	3	5	
$u_2$	4		5	4	4
$u_3$		3		4	3
$u_4$	5		4		5
$u_5$		2		4	

Tabela 1: Matriz Usuário X Item de avaliações

Pela Tabela 1, podemos ver as avaliações dos usuários  $u$  para os itens  $i$ . Nos espaços em branco, não houve avaliação. O objetivo dos Sistemas de Recomendação é inferir qual nota o usuário daria para o item não avaliado, para que ele seja recomendado ao usuário segundo o seu interesse.

Uma das formas de fazer essa inferência é usando técnicas Baseadas em Memória, como o Baseado em Vizinhança. Nessa técnica, é calculado o peso ou a similaridade entre os usuários ou itens e, para produzir a recomendação, é usada a média ponderada dos itens avaliados pela similaridade dos itens.

De acordo com Sarwar [6], para sistemas em que existem muito mais itens do que usuários, é recomendado basear o Sistema de Recomendação em Itens, pois achar usuários correlacionados tornaria a tarefa muito complexa pela esparsidade da matriz  $\mathcal{U} \times \mathcal{J}$ .

Dessa forma, para obter os locais relacionados, pode-se usar cálculos de similaridade, sendo o mais comum a Correlação de Pearson:

$$w_{i,j} = \frac{\sum_{u \in \mathcal{U}} (r_{u,i} - \bar{r}_i) \cdot (r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in \mathcal{U}} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in \mathcal{U}} (r_{u,j} - \bar{r}_j)^2}} \quad (2)$$

sendo que  $\mathcal{U}$  é o conjunto de todos os usuários  $u$  que coavaliaram os itens  $i$  e  $j$ ,  $r_{u,i}$  a avaliação que o usuário  $u$  deu ao item  $i$  (valor numérico de 1 a 5) e  $\bar{r}_i$  a média das avaliações para o item  $i$ .

Com um histórico de itens que o usuário já avaliou, é feita uma soma ponderada entre as similaridades do local sugerido  $k$  com as avaliações de outros locais que o usuário já avaliou:

$$p_{colaborativo}(u, k) = \bar{r}_k + \frac{\sum_{i \in \mathcal{J}} (r_{u,i} - \bar{r}_i) \cdot w_{i,k}}{\sum_{i \in \mathcal{J}} |w_{i,k}|} \quad (3)$$

sendo que  $p_{colaborativo}(u, k)$  é a avaliação estimada do usuário  $u$  para o item  $i$ . Fazendo isso para os locais relevantes, basta ordenar para que se tenha a lista dos *Top N* locais para serem sugeridos ao usuário.

Pelos itens da recomendação serem locais geográficos, o sistema de recomendação sofre o problema do contexto do usuário; ou seja, as recomendações só terão sentido se conhecido onde o usuário está geograficamente. Dessa forma é necessário um Sistema de Recomendação Híbrido, que mescla abordagem colaborativa com abordagem baseada em contexto

Segundo Burke [10], o Sistema de Recomendação com o contexto combinado com o resultado do algoritmo colaborativo forma um Sistema de Recomendação Híbrido com característica de cascata; ou seja, o resultado do algoritmo colaborativo é alterado pelo fator de contexto para se obter o resultado final. Assim, a recomendação final é feita multiplicando o resultado do algoritmo colaborativo com o fator de contexto do usuário  $u$  e o local  $k$ :

$$p_{final}(u, k) = p_{colaborativo}(u, k) \times p_{contexto}(u, k) \quad (4)$$

Vários trabalhos fazem sugestões de locais [7][8] em que o usuário tem um dispositivo móvel que fornece a captura da sua coordenada GPS no momento de gerar a sugestão. No caso do site Apontador e de outros sites similares, muitos usuários apenas visitam as páginas dos locais, não sendo possível descobrir a sua localização exata.

Dessa maneira, espera-se que os locais sugeridos para o usuário pelo algoritmo colaborativo se tornem mais relevantes de acordo com o resultado do fator de contexto.

### III. SISTEMA DE RECOMENDAÇÃO COM CONTEXTO DE LOCAIS

Ao navegar pela busca de locais, o usuário está dando algumas dicas de como é o seu contexto ou região onde ele está, pois irá buscar por locais onde ele deseja frequentar e irá avaliar locais onde ele frequentou. Além disso, ele poderá também fazer *checkin* com a aplicação mobile do Apontador, indicando, no momento, que estava naquele local.

Diferente do contexto de aplicações como *Foursquare*, em que sabemos onde o usuário está no momento por meio dos *checkins*, a maioria dos usuários do site Apontador apenas visita as páginas dos locais nos seu computador pessoal, o que permite visitar as páginas dos locais das mais variadas cidades e distâncias do seu local de residência. Com esse problema, é necessário inferir o contexto do usuário para que a recomendação seja correta.

Com o contexto do usuário definido, deseja-se inferir o quanto o local sugerido  $i$  pelo algoritmo colaborativo está próximo do contexto do usuário. Como locais têm uma coordenada geográfica ( $X, Y$ ) para identificação espacial, os modelos de predição podem ser usados para avaliar se o local está inserido no contexto do usuário. Desta forma, seria possível inferir se o usuário iria ao local sugerido  $i$  se ele visitou recentemente uma lista de locais  $L = \{j_1, j_2, \dots, j_k\}$ .

O modelo proposto por Ye [9] se encaixa nesta necessidade. Para isso, foi definida uma função PL de probabilidade do usuário ir ao local  $i$  pela lista de lugares  $L$  que ele já visitou ou avaliou:

$$p_{contexto} = PL(i, L) = \prod_{j \in L} P(i, j), \quad (5)$$

sendo que  $P$  calcula a probabilidade de um usuário visitar os dois locais apenas pela distância entre eles:

$$P(i, j) = a \times b^{d(i, j)} \quad (6)$$

Ye [9] propõe que a probabilidade de um usuário visitar um local  $i$  e um local  $j$  segue uma distribuição de probabilidade na forma de potência, sendo que  $a$  e  $b$  são os parâmetros da distribuição. Os parâmetros  $a$  e  $b$  são calculados para cada sistema usando método do Gradiente Descendente e conjunto de dados é cada distância entre lugares  $i$  e  $j$  que um mesmo usuário avaliou. A distância usada neste modelo é a distância euclidiana, já que temos as coordenadas  $(X, Y)$  dos locais  $i$  e  $j$ :

$$d(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (7)$$

Este modelo atende apenas quando analisamos o contexto cotidiano do usuário, que seriam os locais que ele vai durante o dia e os lugares próximos à residência do usuário. Por isso, este modelo não sabe distinguir se um novo local muito distante dos outros que o usuário visitou é uma busca fora do contexto do usuário ou se ele está mudando o seu contexto, como uma mudança de cidade ou interesse em viagem de férias.

#### IV. TECNOLOGIA ADAPTATIVA

A área de Tecnologia Adaptativa [11] tem mostrado um forte crescimento com aplicação de técnicas de aprendizado a partir de experiências anteriores e com aplicações na área de Data Mining [12].

O dispositivo adaptativo é composto de uma camada subjacente; que contém um dispositivo não adaptativo com um formalismo já conhecido, que forma o núcleo do sistema; e uma camada adaptativa. Suas ações modificam o núcleo apenas com as funções pré definidas e com os dados de entrada, tornando o dispositivo automodificável.

Assim, um dispositivo adaptativo pode ser formalizado como  $DA = (ND_0, AM)$ , com:

- $DA$ : dispositivo adaptativo;
- $ND_0$ : dispositivo não adaptativo subjacente;
- $AM$ : Mecanismo Adaptativo, tal que  $MA \subset AA \times RDN \times AP$ ;
- $RDN$ : regras do dispositivo não adaptativo;
- $AA$ : conjunto das funções adaptativas anteriores; e
- $AP$ : conjunto das funções adaptativas posteriores.

As funções adaptativas contêm três regras básicas sobre o dispositivo subjacente:

1. Consulta de regra;
2. Adição de regra; e
3. Remoção de regra.

Dispositivos adaptativos são interessantes no ponto que o dispositivo subjacente pode ser usado tanto para processamento quanto para memória. Isso permite fazer tratamentos mais elaborados e se automodificar de acordo com o seu contexto. Podemos ver a adaptatividade como transições no espaço de todos os possíveis dispositivos não adaptativos,

em que as funções adaptativas Posterior e Anterior escolhem qual dispositivo não adaptativo será executado.

#### V. MODELO ADAPTATIVO PROPOSTO

Para melhorar o fator de contexto na recomendação de locais é proposto um modelo adaptativo no dispositivo do contexto dos usuários.

A melhoria se daria em que o fator de contexto não trabalhe apenas com um contexto único como é apresentado na seção III, mas sim com vários contextos ao mesmo tempo, sendo-os inseridos durante a captura de locais que o usuário visitou ou avaliou. A adaptatividade é um conceito que se encaixa bem nessa necessidade, pois a atualização do contexto seria feita online de acordo com a navegação do usuário pelo site e pela busca de locais. Destas formas, poderíamos instantaneamente fazer recomendações com os novos contextos.

Cada contexto é uma lista  $L_n = \{j_1, j_2, \dots, j_k\}$  com os locais que o usuário visitou, tendo tamanho máximo  $\|L\| = K$ , sendo  $K$  um valor inteiro arbitrário.

Um contexto pode ser definido como um local  $i$  dentro de  $L_n$ , cuja distância euclidiana entre  $i$  e os itens  $j_n$  é menor do que um valor  $\theta$  arbitrário para pelo menos  $\frac{\|L\|}{2}$  locais do mesmo contexto. Ou seja, todos os locais estarão dentro de uma região física, sendo que a distância euclidiana é a distância definida entre os locais  $i$  e  $j$  pelas coordenadas  $(X, Y)$  de cada um.

Ao respeitar essa regra, não há nenhuma ação adaptativa, apenas o local mais antigo que o usuário interagiu será retirado na lista se  $\|L\| = K$ , e o novo local  $i$  será adicionado. Nesse momento, o dispositivo adaptativo apenas faz as ações de consulta em vez de executar alguma ação sobre o dispositivo subjacente, que apenas inclui o novo item na sua lista, conforme apresentado na Figura 1 e Figura 2.

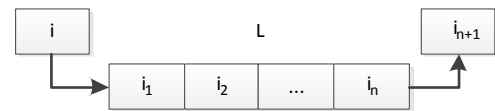


Figura 1 - Inserção de item na lista

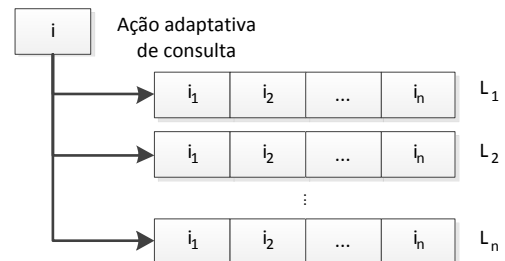


Figura 2 - Ação adaptativa de consulta

Se novo local  $i$  tiver mais do que  $\frac{\|L\|}{2}$  locais distantes pelo limite  $\theta$  em um contexto, será comparado com outro contexto. Se em nenhum contexto o local  $i$  puder ser inserido, então um novo contexto  $L_{n+1}$  será criado. Além disso, os locais dos outros contextos (tal que  $D(i, j) < \theta$ ) serão inseridos no contexto recém criado, respeitando o tamanho  $\|L_{n+1}\| \leq K$ . Para isso serão inseridos os lugares mais próximos de  $i$ .

Nesse momento, a ação adaptativa cria uma nova regra que será o contexto  $L_{n+1}$ , fazendo a inserção do local  $i$  e dos outros locais, tal que  $D(i, j) < \theta$ .

Para a recomendação de um local com múltiplas listas, o valor do algoritmo colaborativo multiplica-se pelo fator de contexto. Havendo múltiplos contextos, o resultado não será apenas uma lista  $TOP N$  de recomendações, mas sim  $n$  listas de recomendações. Para o caso de um sistema que tiver apenas uma lista  $Top N$ , é necessário que o contexto tenha apenas mais importância para aquele local  $i$ :

$$\text{Fator de Contexto} = \arg \max_n P(L_n, i) \quad (8)$$

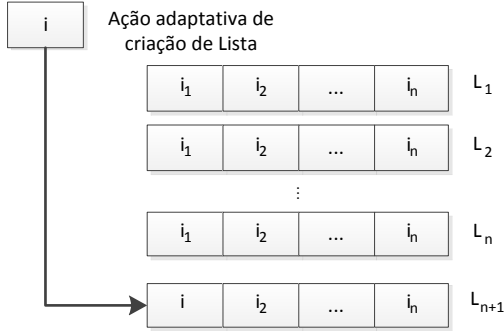


Figura 3 - Ação adaptativa de inclusão

Para fazer a recomendação final ao usuário, basta multiplicar o Fator de Contexto pelo resultado do algoritmo colaborativo. Assim, os locais mais distantes do contexto perderão importância na sugestão e os mais próximos ficarão em destaque.

Como os contextos podem ser temporários, como por exemplo uma viagem de férias, eles podem ser apagados depois de um tempo  $T$ . Dessa forma, se o último local inserido no contexto for inserido a mais tempo do que  $T$ , ele será eliminado, evitando a poluição de dados antigos.

## VI. RESULTADOS ESPERADOS

Para avaliar a melhoria da inclusão da técnica adaptativa no Sistema de Recomendação baseado em Contexto, este será comparado com o mesmo Sistema de Recomendação sem o contexto.

A forma mais tradicional de medir a qualidade de um Sistema de Recomendação de acordo com [13], é pela acurácia da predição. Uma das medidas que será usada é a Raiz do Erro Quadrático Médio ( $REQM$ ), que foi usada na avaliação do prêmio oferecido pela Netflix [14]:

$$REQM = \sqrt{\frac{1}{n} \sum_{u \in \mathcal{U}, i \in \mathcal{I}} (p_{u,i} - r_{u,i})^2} \quad (9)$$

Em que  $n$  é o total de avaliações sobre todos os usuários,  $p_{u,i}$  é a avaliação inferida do item  $i$  ao usuário  $u$ , e  $r_{u,i}$  é a avaliação real.

Outras medidas usadas são mais conhecidas na área de Recuperação de Informação, como a Precisão e o *Recall* [15]:

- **Precisão:** fração dos itens retornados que são relevantes;
- **Recall:** fração dos itens relevantes que são retornados.

## VII. CONCLUSÕES

Esse trabalho propõe a inclusão de tecnologia adaptativa para a recomendação de locais. A técnica estende o conceito de contextos para que seja possível detectar mais de um contexto ao mesmo tempo em que o usuário tenha interesse.

A aplicação desse trabalho não serve apenas para a recomendação de locais, mas também pode ser aplicada para recomendação de outros itens, como venda de produtos no *e-commerce*, em que o usuário pode ter vários interesses distintos na compra ou pode estar comprando para outras pessoas.

## REFERÊNCIAS

- [1] Linden, G., Smith, B., and York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing* 4, 1 (January 2003).
- [2] X. Su and T. M. Khoshgoftaar, A Survey of Collaborative Filtering Techniques, *Advances in Artificial Intelligence* (2009).
- [3] Adomavicius, G., Tuzhilin A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, (2005) 17(6): p. 734-749
- [4] Baeza-Yates, R. AND Ribiero-Neto, B. 1999. *Modern Information Retrieval*. Addison-Wesley Longman, Boston, Mass.
- [5] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35, 12 (December 1992), 61-70.
- [6] Sarwar, B. M., Karypis, G., Konstan, J. A., AND RIEDL, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International World Wide Web Conference (WWW10)*.
- [7] C. Stiller, F. Roß, and C. Ament. Towards Spatial Awareness in Recommender Systems. In *Proceedings of the 4th International Conference for Internet Technology and Secured Transactions - ICITST'09*, Nov. 2009.
- [8] M. Brunato and R. Battiti, "Pilgrim: A Location Broker and Mobility-Aware Recommendation System", Technical Report DIT-02-092, Informatica e Telecomunicazioni, University of Trento.
- [9] M. Ye, P. Yin, W.-C. Lee, and D. L. Lee. Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2011.
- [10] Burke, R., *Hybrid Recommender Systems: Survey and Experiments*, User Modeling and User-Adapted Interaction, v.12 n.4, p.331-370, November 2002
- [11] NETO, J. J., *Adaptive Rule-Driven Devices - General Formulation and Case Study*. Lecture Notes in Computer Science. Watson, B.W. and Wood, D. (Eds.): Implementation and Application of Automata 6th International Conference, CIAA 2001, Vol. 2494,
- [12] Camargo, R., Raunheite, L., *Convolução Adaptativa*. Workshop de Tecnologia Adaptativa (5 : 2011 : São Paulo) Memórias do WTA 2011. - São Paulo ; EPUSP, 2011. 111 p
- [13] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22 (1). 5-53
- [14] Netflix prize, <http://www.netflixprize.com/>
- [15] Ricci, F., Rokach, L., and Shapira, B., editors (2010). *Recommender Systems Handbook*. Springer.



**Celso Vital Crivelaro** é nascido em São Paulo - SP, em 06 de Janeiro de 1985 e graduou-se em Engenharia de Computação pela Escola Politécnica da Universidade de São Paulo (2008). Desde 2009 é mestrando em Engenharia de Computação pela EPUSP. Atualmente é responsável pela área de Recomendações na empresa Apontador ([www.apontador.com.br](http://www.apontador.com.br)) e é membro do IEEE (Institute of Electrical and Electronics Engineers).



**Fabrício J. Barth** é nascido em Ibirama – SC, em 21 de junho de 1978. Graduou-se em Ciências da Computação pela Universidade Regional de Blumenau. É Mestre e Doutor em Engenharia de Computação pela EPUSP (2003 e 2009, respectivamente). Suas áreas de atuação incluem Recuperação de Informação, Aprendizagem de Máquina e Engenharia de Conhecimento. É membro da ACM (Association for Computing Machinery) e da SBC (Sociedade Brasileira de Computação).

**Ricardo Luis de Azevedo da Rocha** é natural do Rio de Janeiro-RJ e nasceu em 29/05/1960. Graduou-se em Engenharia Elétrica na modalidade Eletrônica na PUC-RJ, em 1982. É Mestre e Doutor em Engenharia de Computação pela EPUSP (1995 e 2000, respectivamente). Suas áreas de atuação incluem Tecnologias Adaptativas e Fundamentos de Computação e Modelos Computacionais. Dr. Rocha é membro da ACM (Association for Computing Machinery), do IEEE (Institute of Electrical and Electronics Engineers) e da SBC (Sociedade Brasileira de Computação).