# Identificação de especialistas e recomendação de novos relacionamentos em ambientes de *microblogging*

Samantha Vrabl <sup>1,3</sup>, Jonice Oliveira <sup>1,2</sup>, Cláudia L. R. Motta<sup>1,3</sup>

- <sup>1</sup> Programa de Pós-Graduação em Informática (PPGI) Universidade Federal do Rio de Janeiro Rio de Janeiro RJ, Brasil
  - <sup>2</sup> Departamento de Ciência da Computação (DCC) / Instituto de Matemática (IM) Universidade Federal do Rio de Janeiro Rio de Janeiro RJ, Brasil
- <sup>3</sup> Núcleo de Computação Eletrônica (NCE) / Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais - Universidade Federal do Rio de Janeiro - Rio de Janeiro -RJ, Brasil

svrabl@hotmail.com, jonice@dcc.ufrj.br, claudiam@nce.ufrj.br

Abstract. Microblogging offers information fast, portable and organic and can be considered as significant sources of learning. This article presents a social recommendation engine, which from the temporal analysis of the messages, traces the profile of interest to users, recommending new relationships in order to assist the learning process.

Resumo. Microblogging oferece informações rápidas, portáveis e orgânicas e podem ser considerados como fontes significativas de aprendizagem. Esse artigo apresenta um mecanismo de recomendação social, que a partir da análise temporal das mensagens, traça o perfil de interesse dos usuários, recomendando novos relacionamentos, de maneira a auxiliar o processo de aprendizagem.

# 1. Introdução

O ato de recomendar livros, filmes, locais ou profissionais (para um atendimento médico, por exemplo) corresponde a uma atividade natural e recorrente em nossas vidas. Na internet, processo similar a este também ocorre. Há sites de compra que oferecem recursos inteligentes de recomendação de produtos, bem como ambientes de redes sociais que promovem a combinação social, isto é, recomendam pessoas focalizando a indicação de amigos em comum ("se A é meu amigo e é também amigo de C, então, C pode ser um provável amigo meu"), que, muitas vezes, não estão em linha com os nossos interesses.

A proposta deste artigo é apresentar uma abordagem de combinação social, ou recomendação de pessoas, baseando-se nas mensagens e no perfil do usuário em um *microblogging*, a fim de recomendar pessoas com nível de conhecimento relevante para auxiliar na aquisição de informação. A heurística do modelo considerou a análise multivariada discriminante para determinar possíveis especialistas, contando com a prototipagem computacional para a extração de dados e cálculo das métricas de análise

de redes sociais. Na seção 2 apresentamos a visão geral do modelo proposto, chamado #twintera!, para a seguir (na seção 3), apresentar resumidamente avaliação realizada e na seção 4 apresentamos a conclusão e trabalhos futuros.

### 2. O Modelo #twintera!

Pretende-se, por meio deste modelo, que o indivíduo receba recomendações de pessoas em um *microblogging*, segundo os temas que mais lhe aprouver e identifique aquelas com quem pode se informar e, assim, promova novas possibilidades de aquisição de conhecimento. O tutorial do protótipo do #twintera! pode ser acessado através do link: <a href="http://youtu.be/haneMIHUEqU">http://youtu.be/haneMIHUEqU</a>

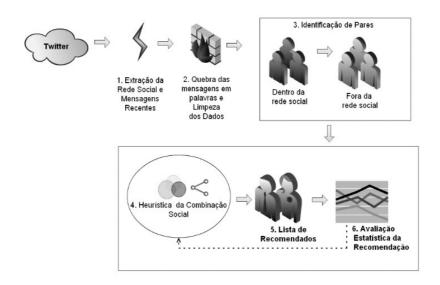


Figura 1. Etapas do Modelo #twintera!

A Figura 1 compreende 6 passos descritos resumidamente a seguir:

- 1. O usuário autoriza o acesso para que o #twintera! possa extrair suas informações do Twitter, tais como nome, amigos, seguidores, listas, biografia, *url*, data de criação no ambiente, localização e última mensagem enviada. O #twintera! recupera as 200 últimas mensagens de cada um dos amigos e seguidores do usuário. São filtradas somente aquelas escritas nos idiomas português e inglês. Não consideramos as mensagens do usuário, mas sim, a de seus amigos e seguidores porque ele declara o interesse que pretende pesquisar no #twintera!. Todos os interesses pesquisados pelo usuário são registrados na base de dados.
- 2. Decomposição das mensagens coletadas dos amigos e seguidores do usuário em palavras e limpeza de termos não-significativos (*stop words*), tais como preposições, artigos, advérbios, números, pronomes e pontuação. Há também a contagem da frequência das palavras usadas. Cada amigo e seguidor do usuário será associado às palavras e à sua frequência. Estas informações (conjunto de palavras e suas respectivas frequências) serão armazenadas na base de dados para futuras consultas.

- 3. Na identificação de Pares, o usuário declara uma palavra de interesse que deseja pesquisar e o #twintera! procura inicialmente, em sua base interna, pessoas que frequentemente comentam sobre tal palavra. A base interna é composta pelos amigos e seguidores (rede social) não só do usuário, mas de todos os outros que já utilizam o #twintera!. Caso a palavra não conste na base interna, uma nova busca ao Twitter (2011) é realizada, na qual o #twintera! utiliza a opção de pesquisa externa padrão do *microblogging*.
- 4.O #twintera! recupera as informações de quantidade de amigos e seguidores que escreveram alguma mensagem que contivesse a palavra procurada de cada um dos pares recomendáveis e monta a rede social do assunto procurado (ou seja, pessoas e interações feitas unicamente sobre o assunto) com base nos pares recomendáveis. São calculadas as métricas de centralidade local ("degree"), proximidade ("geodesic distance") e grau de intermediação ("betweeness") da rede social única. Em seguida, para cada par recomendável, o #twintera! aplica os valores de seguidores, amigos, centralidade local, proximidade e intermediação através de equações matemáticas detalhadamente descrito em Vrabl (2011) e identifica quem são especialistas e não-especialistas para a palavra-chave determinada pelo usuário.
- 5. O #twintera! apresenta os pares recomendáveis em formato de lista na seguinte sequência: primeiro são apresentados os especialistas, em seguida, os que tendem a ser especialistas, os que tendem a não ser especialistas, e os não especialistas. Em cada grupo desses há a ordenação por maior frequência de palavras e maiores valores de centralidade, proximidade e intermediação
- 6. Na avaliação estatística da recomendação há o registro das buscas efetuadas pelo usuário do #twintera! e a comparação entre quantidade de recomendações sugeridas e aceitas pelo usuário.

## 3. Eficácia dos Resultados da Recomendação

Utilizando o #twintera!, identificamos 4 pessoas para cada um dos três interesses: colaboração, web e tecnologia. Nesta etapa, utilizamos uma amostra que avaliou a recomendação feita. Participaram 32 usuários do Twitter, localizados no Brasil em sua totalidade, com equilíbrio entre gêneros de 50% e de formação (Pós-Gradução *Lato Sensu* 41%, *Stricto Sensu* 37%). A atuação profissional é diversificada, com a predominância da área de educação (40%), análise de sistemas (19%) e estudantes universitários (13%). O convite foi feito por *e-mail*, pelo Twitter e Facebook (2011).

Em nossa pesquisa, apresentamos o nome e identificador do Twitter para que os respondentes pudessem analisar as pessoas recomendadas. Temos que 70% dos participantes aceitaram de uma a três recomendações oferecidas, para fins de aquisição de conhecimento (7% acataram somente uma recomendação, 44% aceitaram duas, 20% concordaram com todas as três). Consideramos esta realidade como satisfatória, levando-se em conta que grande parte da rejeição às recomendações (30%) foi por não conhecer as pessoas e não haver pesquisado sobre o tema, embora esta possibilidade fosse informada no questionário.

Acreditamos que houve uma percepção positiva (60%) da importância da recomendação de pessoas em ambientes como o Twitter (2011), quando se considera perfil e análise de redes sociais.

## 4. Conclusão e Propostas Futuras

O #twintera! corresponde a um modelo de combinação social resultante de um conjunto de etapas que se inicia na extração dos dados do usuário no Twitter (2011), segundo uma palavra de interesse e na análise de seu grau de importância na rede social (quantidade de conexões, centralidade, aproximação e intermediação com outros usuários). Em seguida, efetua um cálculo heurístico de discriminância visando a identificar especialistas.

Esta pesquisa aponta para o desenvolvimento de novas aplicações e realização de novos estudos de caso, podendo ser ampliadas em perspectivas, como: a integração da heurística da combinação social e da avaliação estatística da recomendação à implementação realizada para um novo estudo de caso; tratamento semântico entre a palavras de interesse, com a utilização de programas de linguagem natural para ofertar mais qualidade ao processo atual, utilização de redes bayesianas para reajustes automáticos dos coeficientes da equação do modelo #twintera!, assim, visando a refinar a classificação do especialista quando uma nova recomendação aceita é integrada na base de dados, entre outros.

Além do #twintera!, há um conjunto de serviços externos que abordam a análise de perfil do usuário no Twitter (2011) para fornecer dados sobre o seu comportamento ou classificá-lo segundo seu grau de influência. Apresentamos também aqueles que recomendam pessoas e itens. Concluímos que o nosso modelo diferencia-se por propor análise de perfil conjugada à recomendação de pessoas, focalizando seus interesses aliados às métricas da rede social. Atualmente, para se encontrar uma tentativa equivalente ao #twintera! é preciso consultar um conjunto de ferramentas de mídia social. Ressaltamos que o #twintera! não vislumbrou integrar-se com outros serviços externos via API para promover uma combinação social integrada, porque eles são "caixas-pretas" que não discriminam a fórmula de classificação de perfis publicamente (só oferecem o produto final, a classificação), além de sofrerem muitas mudanças ou até tornarem-se inativas. Outro aspecto importante é que se os serviços externos de análise de perfil fossem conjugados com recomendação de pessoas ainda assim não seriam similares ao #twintera!, porque possuem regras diferentes que não consideram análise de redes sociais e heurística similar ao nosso modelo proposto.

### Referências

Facebook (2011), http://www.Facebook.com/, Abril.

Twitter (2011), http://twitter.com/, Abril.

Vrabl, S. (2011) "#twintera!: um modelo de combinação social usando *microblogging*", Dissertação de Mestrado, Programa de Pós-Graduação em Informática, UFRJ, 2011.