

Identificando Colegas para Ajudar em Minhas Dúvidas: Um Estudo Empírico em Comunidades de Perguntas e Respostas

Thiago B. Procaci, Leila C. V. de Andrade, Sean W. M. Siqueira

Programa de Pós-Graduação em Informática – Universidade Federal do Estado do Rio de Janeiro (UNIRIO) – Rio de Janeiro – RJ – Brasil

{thiago.procaci, leila, sean}@uniriotec.br

Abstract. Online communities of questions and answers have become important places for users to exchange information and build knowledge through interactions. People ask questions and want to get answers. Colleagues (other users), called reliable users, who are willing to help and may provide good answers in such communities can support the knowledge building. Therefore, we investigated users' attributes with artificial neural network and clustering algorithm for finding reliable users. The results show that the usage of an artificial neural network is a good approach as around 90% of the users were correctly identified while the clustering algorithm makes easier to find groups of reliable users.

Resumo. As comunidades online de perguntas e respostas tornaram-se lugares para usuários construírem novos conhecimentos. As perguntas são feitas e os usuários esperam obter respostas. Colegas (outros usuários) que possam ajudar a prover boas respostas, chamados de usuários confiáveis, podem apoiar esta construção de conhecimento. Assim, investigamos atributos de usuários de comunidades online juntamente com o uso de rede neural artificial e algoritmo de agrupamento para encontrar os usuários confiáveis das comunidades. 90% dos usuários foram corretamente identificados como confiáveis através de rede neural e o algoritmo de agrupamento possibilitou encontrar grupos de usuários confiáveis com mais facilidade.

1. Introdução

Procurar por soluções ou respostas para problemas pode não ser uma tarefa trivial. Algumas pessoas realizam buscas na Web ou contatam especialistas, porém, em algumas situações, a melhor opção é utilizar as comunidades online de perguntas e respostas para obter a informação desejada e aprender sobre algo. Contudo, nessas comunidades, pessoas podem não receber respostas ou mesmo receber respostas erradas. Uma das alternativas para minimizar tais problemas é identificar pessoas que estão dispostas a ajudar e que são capazes de fornecer boas respostas que, neste trabalho, são chamadas de usuários confiáveis. A motivação principal para encontrar os usuários confiáveis é pautada na importância de trocas de experiências e informações no processo de ensino e aprendizagem. Sob essa perspectiva, um usuário confiável é um colega que pode apoiar os demais na construção do conhecimento (ou um potencial monitor), em uma comunidade online. Neste cenário, este trabalho objetivou buscar por soluções que permitam identificar os usuários confiáveis de comunidades.

Devido às crescentes demandas por conhecimento dentro das organizações e uma disponibilidade limitada de recursos e competências para suprir tais demandas,

muitos profissionais, tanto da indústria quanto da academia, acabam buscando por conhecimento em fontes externas para resolver os seus problemas [Prates *et al.*, 2013], [West *et al.*, 2014]. Essas fontes externas são muitas vezes os motores de busca da Web, sites ou mesmo comunidades online, onde pessoas visam encontrar soluções para seus problemas diários. Alan *et al.* (2013) afirmam que as comunidades online destinadas à aprendizagem são lugares eficientes para se procurar ajuda, pois, em geral, são compostas por indivíduos que compartilham interesses comuns e voluntariamente trabalham para expandir a sua compreensão sobre um domínio do conhecimento. Em geral, os membros dessas comunidades não se conhecem, podem ser identificados por pseudônimos e estão dispostos a ajudar uns aos outros por diversas razões: altruísmo, reputação, a reciprocidade esperada e os benefícios da aprendizagem [Lin e Huang, 2013]. Essas comunidades são fortemente dependentes de seus membros cooperantes. São através dos membros e de suas participações que a comunidade cresce e, como consequência, maiores são as chances de colaborações bem-sucedidas e construções de conhecimentos. Assim, as formas de ensino e aprendizado de hoje não se limitam aos moldes tradicionais, pois, é perfeitamente possível aprender em ambientes mais informais como as comunidades online [Carvalho *et al.*, 2005], [Marcon *et al.*, 2012].

O processo de postagem de perguntas em uma comunidade online e espera por respostas é conhecido como *social query* [Souza *et al.*, 2013], [Banerjee e Basu, 2008]. Esse processo pode ser visto como uma alternativa aos motores de busca. Segundo Horowitz *et al.* (2010), alguns problemas são melhores resolvidos por pessoas, por exemplo, perguntas muito contextualizadas, pedidos de recomendação, pedidos de opiniões, conselhos etc. Huberman *et al.* (2013) afirmam que ambientes que permitem a formação comunidade online com muitos usuários (milhares de usuários no mínimo), como o *Twitter* e o *Facebook*, são lugares bons e eficientes para encontrar informações através do uso de *social query*. Isso se deve à presença de muitos usuários que, por sua vez, aumentam as chances de se receber algum tipo de informação ou resposta.

Todavia, o uso de *social query* também tem suas limitações. Quando uma pergunta é postada em uma comunidade, alguns resultados não esperados podem ser encontrados como: receber respostas erradas ou contraditórias; continuar recebendo respostas mesmo depois de o problema ser resolvido; nunca receber uma resposta, uma vez que, algumas comunidades tendem a priorizar a visualização das postagens mais recentes [Paul *et al.*, 2010]. Diante desse cenário e visando buscar caminhos para minimizar algumas das limitações da *social query* (respostas erradas e ausência de respostas), esse trabalho buscou investigar atributos (métricas) relacionados a usuários de (cinco) comunidades online reais, com o objetivo de encontrar evidências que permitem inferir quais são os usuários mais confiáveis. Além disto, um modelo e um algoritmo de aprendizado de máquina (uma rede neural artificial e um algoritmo de agrupamento) foram utilizados com os atributos dos usuários para encontrar os confiáveis das comunidades. Em outras palavras, esse trabalho buscou combinar as características dos usuários investigadas (atributos ou evidências de usuários confiáveis) com duas abordagens de aprendizado de máquina para que, no final, seja possível identificar usuários confiáveis, com uma precisão aceitável.

Este trabalho está organizado da seguinte forma: a seção 2 apresenta trabalhos relacionados; na seção 3 é apresentado um estudo empírico em cinco comunidades online distintas visando caracterizá-las e apresentar a proposta do trabalho. A seção 4 é destinada à conclusão.

2. Trabalhos Relacionados

Uma alternativa aos motores de busca para a resolução de problemas ou dúvidas são as comunidades online de perguntas e respostas como o *Quora*¹ e o *Yahoo! Answers*², onde os usuários perguntam e respondem de forma voluntária e, após esse processo, acabam aprendendo. Teevan *et al.* (2010) apresentaram um estudo realizado internamente na Microsoft, utilizando suas próprias ferramentas de comunicação, e concluiu que 93,5% dos usuários tiveram suas perguntas respondidas e, em 90,1% dos casos, os usuários obtiveram respostas em menos de um dia.

Paul *et al.* (2010) fizeram estudos similares no Twitter e concluíram que somente 18,7% das perguntas postadas por um usuário do Twitter recebiam respostas. Também concluíram que o número de respostas recebidas por um usuário tem uma correlação positiva com o seu número de seguidores. Além disso, 67% das perguntas respondidas no Twitter obtinham respostas de modo relativamente rápido (em menos de 30 minutos). Uma das explicações da baixa porcentagem de respostas recebidas se deve ao fato Twitter priorizar a visualização de postagens mais recentes. Logo, é provável que alguns seguidores nem fiquem sabendo da existência de uma determinada pergunta.

Estudos para encontrar os usuários confiáveis em comunidades foram explorados em outros trabalhos científicos. Alguns utilizaram técnicas de recuperação de informações com processamento de linguagem natural para identificar as competências de um usuário [Krulwich e Burkey, 1996] onde, geralmente, os textos produzidos no ambiente virtual são representados através de um vetor de termos (*tokens*) com a sua respectiva frequência. Desta forma, é possível inferir qual o tipo de competência que cada usuário tem baseado em seus discursos. Todavia, o uso da abordagem com foco em recuperação de informação torna difícil elencar o nível de competência de cada usuário, uma vez que, é difícil julgar se um usuário fornece uma boa resposta somente fazendo um parser de seus textos produzidos na comunidade e, em seguida, processando-os [Zhang *et al.*, 2007]. Balog *et al.* (2009) propuseram uma forma para identificar os usuários confiáveis baseada em consultas feitas em um ambiente e uma coleção de textos associados aos candidatos a confiáveis. Este trabalho baseado em técnicas de recuperação de informações e métodos probabilísticos visava determinar a relevância entre uma consulta e os candidatos a confiáveis. Outro trabalho similar foi proposto por Liu *et al.* (2012), em que foi proposto um *framework* que gerava automaticamente os perfis especializados dos usuários da comunidade. Esses perfis continham informações sobre as competências dos usuários e eram construídos baseados na associação entre os tópicos da comunidade com o perfil comum do usuário.

Outra abordagem utilizada é baseada em algoritmos de ranqueamento em grafos para encontrar os usuários confiáveis de uma rede. A ideia dessa abordagem é aplicar algoritmos na comunidade (representada através de um grafo) que atribui um número para cada usuário simbolizando seu grau de competência em algum assunto. Campbell *et al.* (2003) utilizaram o algoritmo de ranqueamento *HITS* em grafos para encontrar os usuários confiáveis que faziam parte de uma lista de e-mail. Esta abordagem baseada em grafos se mostrou eficiente, mas estes estudos tinham uma fraqueza: o tamanho pequeno da rede analisada e os resultados podiam não refletir a realidade. Zhang *et al.* (2007) propuseram a construção de um algoritmo baseado em grafos (adaptação do

¹ Site: <http://www.quora.com/>

² Site: <https://answers.yahoo.com/>

algoritmo *Page Rank*) e analisaram métricas (número de respostas, por exemplo) para o mesmo fim, porém, aplicadas em um fórum de discussão online tradicional. Apesar da abordagem de Zhang *et al.* (2007) ter se mostrado interessante, os autores concluíram, através de simulações, que comunidades com diferentes características devem ser analisadas separadamente, pois as características podem influenciar nos resultados obtidos, sendo necessárias adaptações nas medidas ou nas técnicas utilizadas. Alan *et al.* (2013) propuseram uma nova forma de identificar os usuários confiáveis, construindo um modelo híbrido da abordagem baseada em recuperação de informações com a baseada em algoritmos de ranqueamento em grafos.

Banerjee e Basu (2008) apresentaram um algoritmo probabilístico que possibilitava direcionar perguntas para os usuários mais aptos a respondê-la. Esse algoritmo funcionava baseado em ações repetidas na rede no passado. Davitz (2007) considerou uma entidade global do sistema (agente) que monitorava a rede e decidia quais usuários receberiam (visualizariam) uma determinada questão postada, através de uma análise probabilística. Todavia, essa solução baseada em agentes foi testada somente em uma comunidade pequena. Souza *et al.* (2013) propuseram um algoritmo para encontrar os usuários confiáveis que faziam parte da lista de seguidores de um usuário do Twitter. A ideia desse trabalho era encontrar o usuário seguidor com o perfil mais adequado para responder a uma pergunta no Twitter e o algoritmo proposto se mostrou eficaz para encontrar os usuários confiáveis no Twitter.

A ideia do trabalho apresentado neste artigo é revisitar a abordagem baseada em grafos com algoritmo de ranqueamento, mesclada com recuperação de informações e com o uso de técnicas de aprendizado de máquina. O trabalho é baseado em grafos e em recuperação de informação pois as comunidades são representadas através de um grafo e serão extraídos metadados (informações do usuário) das comunidades para realização de análises. Além disso, também é baseado em aprendizado de máquina porque são utilizados uma rede neural artificial e um algoritmo de agrupamento com a finalidade de encontrar os usuários de acordo com o seu grau de confiabilidade. Todavia, este trabalho propõe uma forma diferenciada para encontrar os usuários confiáveis, denominada de “análise por partes”. Essa análise consiste em dividir uma comunidade em vários componentes (partes) e analisá-los separadamente visando investigar lugares na rede onde os usuários confiáveis mais interagem ou participam. A partir dessa análise, é possível classificar um usuário, de acordo com o seu grau de confiabilidade, através de uma rede neural artificial ou um algoritmo de agrupamento.

3. Estudo Empírico

Com a finalidade de testar a proposta desse trabalho, foi necessário extrair um conjunto de dados de comunidades online reais. Nessas comunidades, assim como outras similares, as discussões têm uma estrutura de trilhas (*threads*), ou seja, um usuário posta uma pergunta e, logo após, outros usuários postam respostas ou comentários relativos à pergunta. Cada *thread* pertence a pelo menos uma categoria da comunidade (ex.: álgebra, cálculo) e cada usuário é avaliado por outros usuários, baseado em suas perguntas ou respostas postadas. Esse esquema de avaliação permite que os usuários construam a sua reputação na rede, podendo ser positiva ou negativa.

Foram escolhidas cinco comunidades distintas de perguntas e respostas: Biology Q&A (BQA - 2.317 usuários, 4.549 *threads* 5.734 respostas e 15.545 comentários), comunidade destinada ao aprendizado de biologia; English Language and Usage (ELU -

20.408 usuários, 30.044 *threads*, 79.978 respostas e 216.893 comentários), comunidade voltada para o aprendizado da língua inglesa; Physics Q&A (PQA - 15.753 usuários, 31.678 *threads*, 51.838 respostas e 166.821 comentários), comunidade destinada ao aprendizado de física; Mathematics Q&A (MQA - 51.245 usuários, 149.948 *threads*, 215.346 respostas e 669.981 comentários), comunidade voltada para o aprendizado de matemática; Travel Answers (TA - 3.579 usuários, 5.529 *threads*; 10.526 respostas; 26.267 comentários), comunidade destinada ao esclarecimento de dúvidas sobre viagens. Para realizar algumas análises necessárias neste trabalho, foi preciso representar as comunidades através de grafos. Zhang *et al.* (2007) propuseram o uso de grafo direcionado para representar esse esquema de perguntas e respostas. Nessa representação, os nós do grafo representam os usuários e as arestas representam as interações entre usuários. Desta forma, se o usuário A posta uma pergunta e o usuário B responde, então o grafo terá um nó A representando o usuário A e um nó B representando o usuário B. Além disso, esse grafo terá uma aresta que sairá do nó A em direção ao B, simbolizando que B respondeu a A. Este trabalho propõe uma pequena extensão desse modelo objetivando representar melhor as interações entre os usuários das comunidades. Como nas comunidades analisadas é possível também comentar uma pergunta ou uma resposta, seguindo a mesma ideia do grafo proposto por Zhang *et al.* (2007), caso um usuário X comente uma pergunta do usuário Y, então uma aresta sairá do usuário Y e chegará no usuário X. Da mesma forma, caso um usuário Z comente uma resposta do usuário K, então uma aresta sairá do usuário K e chegará ao usuário Z.

3.1. Atributos dos Usuários

Visando analisar métricas que possam indicar se um usuário é confiável, foram elencados alguns atributos e uma métrica denominada Índice de Confiança foi proposta. As métricas analisadas e comparadas com a métrica proposta foram: **Entropia do usuário**: a entropia é uma medida que, no contexto desse trabalho, tem como objetivo estudar o foco de um usuário em determinados assuntos (categorias) na comunidade; **Número de respostas e número de comentários**: acredita-se que a reputação de um usuário é construída através de suas boas respostas e seus bons comentários; **z-score**: a ideia dessa medida é combinar o número de perguntas com o número de respostas de um usuário. Responder a muitas questões pode ser um indício que um usuário é confiável, mas perguntar muito pode ser um indício que esse mesmo não seja confiável. A ideia dessa medida é buscar o equilíbrio entre o número de perguntas e respostas do usuário; **Grau de entrada**: como as comunidades foram representadas através de grafos, o grau de entrada significa o número de pessoas que um usuário respondeu. Acredita-se que a reputação de um usuário tem relação com a quantidade de pessoas que ele responde; **Page Rank**: foi selecionado um algoritmo de ranqueamento visando averiguar se seu uso pode encontrar os usuários confiáveis em uma rede; **Índice de Confiança**: métrica proposta neste trabalho para combinar o grau de participação de um usuário, o seu foco em determinados assuntos da comunidade e também o quão recente são as participações.

Essas medidas foram calculadas e correlacionadas estatisticamente com a reputação de cada usuário, fornecida pelas próprias comunidades (Tabela 1). Essa reputação é oriunda do esquema de avaliações de perguntas e respostas dos usuários presentes nas comunidades. Na Tabela 1, “#Resp” se refere ao número de resposta, “#Com” ao número de comentários, “R+C” à soma do número de respostas com

comentários, “G. entrada” ao grau de entrada e “Ind. Conf.” à índice de confiança. Analisando a Tabela 1, é possível verificar que as correlações obtidas foram fortes (acima de 0,7) para “#Resp”, “#Com”, “R+C”, “z-score”, “G. entrada”, *Page Rank* e “Ind. Conf.”. Isto significa que esses atributos podem ser fortes indícios de que um usuário é confiável. Ou seja, quanto maior for o valor desses atributos, maior a chance de um usuário ser confiável. Já que a entropia se correlaciona moderadamente (valores entre 0,3 e 0,7) com a reputação do usuário. Diante disso, se pode concluir que um usuário com mais alta entropia (menos foco em alguns assuntos), pode ser um indicador moderado de que este tem alta reputação. Ou seja, um usuário que participa de várias categorias da comunidade tem maiores chances de ser confiável. Mais detalhes podem ser encontrados em (Procaci *et al.*, 2014a), que corroboram os resultados apresentados em (Procaci *et al.*, 2014b), que realizou um estudo similar em um grupo de discussões do *Facebook* sobre Java.

Tabela 1. Coeficiente de Correlação de Pearson (atributos vs reputação)

Comunidade	#Resp	#Com	R + C	z-score	G. entrada	Page Rank	Entropia	Ind. Conf.
BQA	0,92	0,84	0,89	0,79	0,91	0,90	0,59	0,90
ELU	0,92	0,76	0,82	0,81	0,88	0,86	0,36	0,92
PQA	0,91	0,73	0,81	0,70	0,82	0,81	0,33	0,89
MQA	0,89	0,88	0,90	0,81	0,90	0,89	0,36	0,88
TA	0,97	0,83	0,91	0,76	0,93	0,91	0,44	0,96

3.2. Particionando Comunidades

Esta seção descreve a divisão das comunidades em partes menores e a verificação de em quais partes são obtidas as melhores correlações dos atributos dos usuários com a reputação. Para dividir as comunidades (representadas através de grafos), foi utilizada a estrutura *Bow Tie* [Broder *et al.*, 2000], que reparte um grafo em seis componentes: *Core*, *IN*, *OUT*, *Tendrils*, *Tubes* e *Disconnected*.

Tabela 2. Estrutura Bow Tie

Comunidade	Core	IN	OUT	Tendrills	Tubes	Disconnected
BQA	47,48%	32,59%	11,52%	2,67%	0,00%	5,74%
ELU	48,03%	25,91%	18,89%	3,10%	0,02%	4,05%
PQA	49,12%	29,81%	13,01%	3,08%	0,01%	4,97%
MQA	50,54%	33,93%	8,48%	1,74%	0,01%	5,30%
TA	41,32%	28,95%	22,60%	3,10%	0,03%	4,00%

Tabela 3. Coef. Correlação de Pearson BQA – Estrutura Bow Tie

Atributo	Geral	Core	IN	OUT	Tendrils
#Resp	0,92	0,93	0,05	-0,14	0,03
#Com	0,84	0,83	0,11	0,43	0,47
R + C	0,89	0,90	0,12	0,34	0,40
z-score	0,79	0,83	0,008	0,36	0,26
G. Entrada	0,91	0,92	-	0,32	0,15
Entropia	0,59	0,58	0,11	0,3	0,25
Ind. Conf.	0,90	0,91	0,03	0,13	0,30

Page Rank	0,90	0,91	-	0,29	0,14
-----------	------	------	---	------	------

No contexto do trabalho, o *Core* contém os usuários que frequentemente se ajudam mutuamente (ex.: ora pergunta, ora responde). O *IN* é aquele que contém usuários que somente fazem perguntas e obtêm respostas ou comentários de algum membro do *Core*. O *OUT* é aquele que contém os usuários que respondem ou comentam mensagens postadas por algum membro do *Core*. Os componentes *Tendrils* e *Tubes* se conectam no componente *IN* ou no componente *OUT*, porém, não se conectam no *Core*. Os *Disconnected* são aqueles que não se enquadram em nenhum dos componentes anteriores. A Tabela 2 mostra as porcentagens de usuários das comunidades em cada componente da estrutura. Através dela, percebe-se que as comunidades analisadas são lugares onde grande parte das pessoas deseja ajudar e ser ajudada devido a porcentagem de usuários no *Core*.

A Tabela 3 mostra as correlações dos atributos dos usuários com a reputação (da comunidade BQA), ora considerando a comunidade como um todo (Geral - apresentado na seção 3.1) ora considerando os componentes da estrutura *Bow Tie*. As melhores correlações, em geral, foram obtidas no *Core*. Isto significa que, para encontrar os usuários confiáveis, talvez seja interessante considerar somente os atributos dos usuários no *Core* em vez de considerar a comunidade como um todo. As correlações das demais comunidades foram omitidas devido a restrições de espaço no artigo, porém, os resultados foram similares ao da Tabela 3. Mais detalhes sobre o particionamento das comunidades e conceitos envolvidos podem ser encontrados em [Procaci, 2014] e [Procaci *et al.*, 2014c].

3.3. Aprendizado de Máquina para Encontrar Usuários

Uma vez estudados os atributos que podem ser evidências que um usuário é confiável, buscou-se uma forma de classificar um usuário em confiável ou não. Para isto, foram estudados dois classificadores (aprendizado de máquina): uma rede neural artificial (*Perceptron* Multicamadas) e um algoritmo agrupamento (*K-means*). Contudo, foi preciso definir as possíveis classificações de um usuário para avaliar os classificadores. Diante disto, foi proposto um esquema de médias aritméticas que criam zonas de classificação dos usuários, variando de não confiável até extremamente confiável. Esse esquema de médias é baseado na reputação fornecida pelas comunidades, considerando-se o maior e o menor valores. A título de exemplo, considere a maior reputação igual a MAX e a menor igual a MIN. Uma vez conhecido a maior e a menor reputação, é calculada a média aritmética entre elas. O resultado desta operação é chamado de M1. Em seguida, é calculada a média aritmética entre M1 e MAX e o resultado é denominado M3. Posteriormente, calcula-se a média aritmética entre M1 e M3 e o resultado é chamado de M2. Depois disso, é calculada a média aritmética entre M3 e MAX e o resultado é chamado de M4. Por fim, é calculada a média aritmética entre M4 e MAX e o resultado é denominado M5. Uma vez conhecendo esses intervalos, um usuário é considerado: não confiável, se reputação \leq M1; pouco confiável, se M1 < reputação \leq M2; razoavelmente confiável, se M2 < reputação \leq M3; confiável, se M3 < reputação \leq M4; muito confiável, se M4 < reputação \leq M5; e extremamente confiável, se M5 < reputação \leq MAX.

Com estas classificações, a rede neural (tendo como entrada as métricas apresentadas na seção 3.1 e a saída os intervalos de classificação do esquema de médias) foi treinada e aplicada nas comunidades. Constatou-se que a rede neural

consegue classificar corretamente cerca de 90% dos usuários utilizando o *Core* inteiro ou 20% dele, de cada comunidade, como dados da fase de treinamento da rede neural. Além disso, a rede neural tem capacidade de generalização, comprovada através da validação cruzada e do coeficiente Kappa. Em síntese, uma vez treinada a rede neural, ela é capaz de classificar usuários de uma comunidade sem a necessidade de um mecanismo de reputação.

Contudo, a rede neural de treinamento ainda depende de algum mecanismo prévio (como a reputação) para classificar alguns usuários (no mínimo 20% do *Core*) de acordo com os intervalos do esquema de médias. O outro método de aprendizado de máquina escolhido para encontrar os usuários confiáveis foi o algoritmo de agrupamento *k-means*. Foram realizados testes utilizando o *k-means* com a finalidade de encontrar grupos similares de usuários. Foi descoberto que quanto maior for o número médio de respostas de um grupo, maiores são as chances de um usuário ser confiável. Dado este fato, uma abordagem para saber o grau de confiabilidade de cada usuário dentro de um grupo seria verificar o número de respostas de cada usuário. É importante ressaltar que este método pode ser realizado independentemente de qualquer mecanismo de reputação ou classificação prévia dos usuários. Mais detalhes podem ser encontrados em [Procaci, 2014] e [Procaci *et al.*, 2014d].

4. Conclusão

Este trabalho teve como finalidade buscar formas de encontrar colegas que possam ajudar os demais em suas dúvidas, ou seja, os usuários confiáveis de comunidades online destinadas à aprendizagem. Observa-se que a educação atual não se limita aos moldes tradicionais, sendo possível educar e aprender em ambientes mais informais, como as comunidades online, graças à evolução tecnológica. Todavia, mesmo que novas formas para apoiar a educação tenham aparecido, a importância de promover interações, trocas de experiências, para a construção de conhecimento continua sendo desejada em um ambiente voltado para o ensino e aprendizagem.

A principal contribuição é um conjunto de procedimentos analisados para encontrar os usuários confiáveis, ou seja, aqueles que estão dispostos a ajudar os demais colegas e em geral dão boas contribuições (sendo potenciais monitores ou quem sabe tutores) em uma comunidade.

Entre os trabalhos futuros, podemos citar a análise temporal da confiabilidade dos usuários, uma vez que usuários podem estar dispostos a ajudar em um determinado período, mas em outro não. Também serão estudadas variações temáticas de confiabilidade, ou seja, neste trabalho foram consideradas comunidades de perguntas e respostas que são temáticas, mas em comunidades em que as discussões envolvem diferentes assuntos ou áreas de conhecimento, usuários confiáveis em um tema podem não ser confiáveis em outro. Assim, considerar estas variações temáticas na detecção das confiabilidades dos usuários também é algo importante.

Finalmente, os procedimentos adotados na identificação de usuários confiáveis nestas comunidades de perguntas e resposta também poderiam ser verificados em um sistema voltado para a educação online, buscando mais confirmações para o estudo apresentado. Deste modo, seria interessante identificar se os usuários confiáveis realmente auxiliam os colegas a resolver suas dúvidas, e como eles podem ser estimulados nesta missão em um ambiente educacional.

Referências

- Ackerman, M.S., Wulf, V., Pipek, V. (2002). (eds.). *Sharing Expertise: Beyond Knowledge Management*. MIT Press, 2002.
- Alan, Wang G., Jian, Jiao, Abrahams, Alan S., Fan, Weiguo. and Zhang, Zhongju. (2013) ExpertRank: A topic-aware expert finding algorithm for online knowledge communities, *Decision Support Systems*, Volume 54, Issue 3, February 2013, Pages 1442-1451, ISSN 0167-9236,
- Balog, K., Azzopardi, L., Rijke, M. D. (2009). A language modeling framework for expert finding. *Information Processing and Management*, 45(1), 1–19.
- Banerjee, A., Basu, S. (2008). A social query model for decentralized search. *Proc. 2nd Workshop on Social Network Mining and Analysis*, ACM Press, 2008.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., TOMKINS, A., WIENER, J. (2000). Graph structure in the Web. *Computer Networks*, 33 (1-6). 309-320.
- Campbell, C.S., Maglio, P.P., Cozzi, A. and Dom, B. (2003). Expertise identification using email communications. In the twelfth international conference on Information and knowledge management, New Orleans, LA, 2003, 528-231.
- Carvalho. R. A., Nevado, R. A, Menezes, C. S. (2005). *Arquiteturas Pedagógicas para Educação a Distância: Concepções e Suporte Telemático*. Anais XVI Simpósio Brasileiro de Informática na Educação. Juiz de Fora – MG.
- Davitz, J., Yu, J., Basu, S. Gutelius D. and Harris, A. (2007). iLink: search and routing in social networks. *Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, 2007, pp. 931-940.
- Horowitz, D., Kamvar, S. (2010). The anatomy of a large-scale social search engine. *Proc. of the 19th International Conference on World Wide Web (WWW)*, ACM Press, 2010, pp. 431-440.
- Huberman, B., Romero D., Wu, F. (2009), Social networks that matter: Twitter under the microscope. *First Monday*, vol. 14, 2009, pp. 1-8.
- Krulwich, B., Burkey, C. (1996). ContactFinder agent: answering bulletin board questions with referrals. In the 13th National Conference on Artificial Intelligence, Portland, OR, 1996, 10-15.
- Lin, F., Huang, H. (2013). Why people share knowledge in virtual communities?: The use of Yahoo! Kimo Knowledge+ as an example. In: *Internet Research*, 23(2). 133-159. <http://dx.doi.org/10.1108/10662241311313295>.
- Liu, X., Wang, G.A., Johri A., Zhou, M. and Fan, W. (2012). Harnessing global expertise: a comparative study of expertise profiling methods for online communities, *Information Systems Frontiers* (2012) 1–13.
- Marcon K., Machado, J. B., Carvalho, M. J. S. (2012) *Arquiteturas Pedagógicas e Redes Sociais: Uma experiência no Facebook*, Simpósio Brasileiro de Informática na Educação, Rio de Janeiro.
- Page, L., Brin, S., Motwani, R. and Winograd. (1998), T. The Pagerank Citation Ranking: Bringing Order to the Web, Stanford Digital Library Technologies Project,

1998.

- Paul, S., Hong L. and Chi, E. (2013). Is twitter a good place for asking questions? a characterization study. Proc. Fifth AAAI International Copyright (c) IARIA, 2013. ISBN: 978-1-61208-280-6 152 ICIW 2013: The Eighth International Conference on Internet and Web Applications and Services Conference on Weblogs and Social Media (ICWSM), 2011, pp. 578-581.
- Prates, J.C., Fritzen, E., Siqueira, S.W.M., Braz, M.H.L.B., Andrade, L.C.V. (2013). Contextual web searches in Facebook using learning materials and discussion messages. In: Computers in Human Behavior, 29(2). 386-394. <http://dx.doi.org/10.1016/j.chb.2012.06.002>.
- Procaci, T. B. (2014). Usuários Confiáveis em Comunidades Online: Um Estudo Empírico Envolvendo Análise de Métricas e Aprendizado de Máquina. Dissertação de Mestrado, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.
- Procaci, T. B., Siqueira, S. W. M., Andrade, L. C. V. (2014a). Encontrando Usuários Confiáveis em Comunidades Online de Perguntas e Respostas Através de seu Índice de Confiança. In: Simpósio Brasileiro de Sistemas de Informação (SBSI), Londrina. Anais. Londrina: SBC, 2014. v. 10. p. 675-686.
- Procaci, T. B., Siqueira, S. W. M., Andrade, L. C. V. (2014b). Finding Experts on Facebook Communities: Who Knows More? In: World Summit on the Knowledge Society (WSKS), Venice. Proceedings of the 7th World Summit on the Knowledge Society. Athens: ORS, 2014. v. 7. p. 1-10.
- Procaci, T. B., Siqueira, S. W. M., Andrade, L. C. V. (2014c). Finding Reliable People in Online Communities of Questions and Answers - Analysis of Metrics and Scope Reduction. In: 16th International Conference on Enterprise Information Systems, Lisbon. Proceedings of the 16th International Conference on Enterprise Information Systems. v. 16. p. 526-535.
- Procaci, T. B., Siqueira, S. W. M., Braz, M.H.L.B., Andrade, L. C. V. (2014d). How to find people who can help to answer a question? - Analyses of metrics and machine learning in online communities. Computers in Human Behavior, v. 47, p. 1.
- Souza, C. C., Magalhães, J. J., Costa, E. B.; Fachine, J. M. (2013). Social Query: A Query Routing System for Twitter. In: The Eighth International Conference on Internet and Web Applications and Services (ICIW), 2013. Roma. Proceedings of the International Conference on Internet and Web Applications and Services.
- Teevan, J., Morris, M. and Panovich, K. (2010). What do people ask their social networks, and why? A survey study of status message Q&A behavior. Proc. 28th International Conference on Human Factors in Computing Systems (CHI), ACM Press, 2010, pp. 1739-1748.
- West, J., Salter, A., Vanhaverbeke, W., Chesbrough, H. (2014). Open innovation: The next decade. In: Research Policy, 43(5). 805-811.
- Zhang, J., Ackerman, M.S., Adamic, L. (2007). Expertise networks in online communities: structure and algorithms, Proceedings of the 16th international conference on World Wide Web, May 08-12, 2007, Banff, Alberta, Canada.