

Uso de Regressão Logística para Modelar e Avaliar a Credibilidade em Aplicações Web

Rafael Lima ¹, Adriano Pereira ¹

¹Departamento de Engenharia de Computação (DECOM)
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Belo Horizonte - MG, Brasil

rafaellima@decom.cefetmg.br, adriano@decom.cefetmg.br

Abstract. *The popularization of Web applications has given rise to new services every day, demanding mechanisms to ensure the credibility of these online services. In this work we use logistic regression to design and evaluate the credibility of a Web application. In order to validate our proposed methodology, we perform experiments using an actual dataset, from which we evaluated different credibility models using distinct types of information sources. The obtained results are very good, showing representative gains, when compared to a baseline. The results show that the proposed methodology are promising and can be used to enforce trust to users of services on the Web.*

Resumo. *A popularização das aplicações Web tem feito surgir novos serviços a cada dia, demandando mecanismos que assegurem a credibilidade desses serviços. Neste trabalho utilizamos regressão logística para modelar e avaliar a credibilidade de um serviço da Web, considerando diferentes critérios associados ao serviço e seus fornecedores. A fim de validar nossa metodologia, executamos experimentos usando uma base de dados real, a partir da qual avaliamos esses modelos de credibilidade. Os resultados obtidos são muito bons, apresentando ganhos representativos, quando comparados à linha-de-base, mostrando assim que a metodologia proposta é promissora e pode ser usada para fortalecer a confiança dos usuários nos serviços providos na Web.*

1. Introdução

A popularização das aplicações da Web 2.0, onde usuários podem interagir cada vez mais, criando e compartilhando uma diversidade de conteúdos, negociando produtos e estabelecendo novas comunidades, representou a maior revolução da forma como usuários e corporações utilizam a Web. Essa revolução trouxe desafios relacionados à credibilidade dos serviços e informações disponíveis na Web. Assim, nesse cenário, mecanismos que auxiliem os usuários a avaliar a credibilidade tem se tornado cada vez mais essencial.

Bibliotecas digitais e mercados eletrônicos são exemplos de aplicações que requerem mecanismos para avaliar a credibilidade. Muitas dessas aplicações já provêm sistemas para lidar com isso, tais como sistemas de reputação.

Modelos de credibilidade diferem dos modelos de reputação, os quais são bastante estudados na literatura [Jøsang et al. 2007, Sabater and Sierra 2005], porque eles não consideram apenas o *feedback* dos usuários, mas também um conjunto de atributos,

que podem ser relacionados ao serviço provido e ao usuário fornecedor do serviço, como uma forma para obter uma avaliação mais completa e efetiva de um dado serviço da Web.

Neste projeto de iniciação científica, utilizamos regressão logística para modelar e avaliar a credibilidade de aplicações da Web [Lima and Pereira 2011]. Essa avaliação é baseada em uma amostra representativa de transações eletrônicas que possuem opinião dos usuários e um *ranking* que representa uma escala de credibilidade gerada pelo modelo. Quanto maior a capacidade dos modelos em posicionar vendedores que oferecem serviços satisfatórios (que são bem qualificados a partir do *feedback*) no topo desse ranking, melhor é a qualidade do modelo. Realizamos experimentos usando dados reais de um mercado eletrônico, com o qual avaliamos os modelos de regressão logística usando diferentes tipos de fonte de informação, tais como atributos relacionados às características das ofertas, *expertise* do vendedor e sua qualificação. Os resultados apresentaram ganhos representativos, quando comparados a modelos adotados como linha-de-base experimental (*baseline*).

O restante deste artigo está organizado da seguinte forma. A seção 2 descreve alguns trabalhos correlatos ao tema de pesquisa deste projeto. A seção 3 apresenta uma breve descrição sobre regressão logística, base de nosso trabalho. A seção 4 descreve nossa metodologia, a qual será aplicada na seção 5 em nosso estudo de caso com dados reais providos pela empresa Universo Online (UOL). Finalmente, a seção 6 apresenta as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Nos últimos anos, o conceito de credibilidade na Web passou a ser mais estudado devido à necessidade de mensurar o quanto um usuário confia em um serviço ou informação disponível nesse ambiente digital.

A credibilidade de aplicações da Web se tornou um assunto multidisciplinar, onde pesquisadores das áreas de comunicação tem focado em uma análise mais qualitativa (e subjetiva) da credibilidade [Flanagin and Metzger 2007], enquanto pesquisadores da área de computação tem focado em métricas mais objetivas. Os métodos propostos na área de ciência da computação são fortemente baseados em confiança (*trust*) e reputação [Guha et al. 2004], além de rankings de credibilidade que levam em conta a fonte da informação [Amin et al. 2009] e o seu conteúdo [Juffinger et al. 2009].

Mecanismos de reputação são baseados em opiniões virtuais, fornecidas por pessoas que geralmente não se conhecem pessoalmente. Portanto, a confiança neste meio eletrônico é mais difícil de ser estabelecida, se comparado ao mundo real. Tomando uma visão ampla, nesses mercados eletrônicos uma reputação do comprador representa a probabilidade do pagamento do bem adquirido e a reputação do vendedor representa a probabilidade da entrega bem sucedida do item anunciado (produto que foi comprado) após o pagamento [Houser and Wooders 2006]. Essas probabilidades estão relacionadas com a confiança [Melnik and Alm 2002].

Mercados eletrônicos a cada dia estão se tornando mais populares. Diversos trabalhos investigam os sistemas de reputação e como eles induzem ao comportamento cooperativo. Dellarocas [Dellarocas 2006] realizou uma profunda revisão sobre este tópico. Ao mesmo tempo que fornece incentivos para um bom comportamento dos

usuários, sistemas de reputação também podem fomentar comportamentos enganosos. b Klos et al. [Klos and Alkemade 2005] analisaram o efeito da confiança e da reputação sobre os lucros obtidos através de intermediários em relações comerciais eletrônicas. Diferentes esquemas de propagação de confiança e desconfiança em comércio eletrônico foram estudados e avaliados em Guha et al. [Guha et al. 2004]. Resnick et al. [Resnick and Zeckhauser 2002] mostrou que os vendedores com maior reputação vendem seus produtos mais facilmente, entretanto os ganhos (preços) são reduzidos.

Os trabalhos aqui brevemente descritos sugerem a crescente necessidade de prover novos modelos de credibilidade que forneçam subsídios a usuários de serviços online, para que assim possam atuar com mais segurança e confiança na Web.

3. Regressão Logística

Regressão logística é uma técnica estatística que produz, a partir de uma série de variáveis explicativas, um modelo que permita a predição de valores tomados por uma variável dependente categórica. Assim através de um modelo de regressão, é possível calcular a probabilidade de ocorrência de um evento, através da função de ligação, conforme descrita na Equação 1:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}}, \quad (1)$$

onde $\pi_i(x)$ é a probabilidade de sucesso quando o valor da variável preditiva é x , β_0 é uma constante usada para ajuste e β_i são os coeficientes das variáveis preditivas [Hosmer 2000].

Para encontrar a estimativa dos coeficientes β na Equação 1, é usada a técnica de máxima verossimilhança que maximiza a probabilidade de obter o grupo observado de dados, através do modelo estimado. Na regressão logística essa técnica pode ser resolvida pelo método de *Newton-Raphson*, o qual gera uma sequência de soluções que convergem para o ponto máximo da função de verossimilhança [Casella and Berger 2002].

O modelo de regressão pode assumir natureza ordinal ou nominal, dependendo dos valores que a variável dependente pode assumir [Agresti 1996]. Neste projeto será usado a regressão logística nominal, pois não há uma ordem entre as categorias das variáveis. Neste trabalho a variável dependente contém duas categorias (variável dicotômica), portanto será usado o modelo de regressão logística binária, com multi-variáveis, ou seja, mais de uma variável independente.

Para a realização da regressão logística, é importante explicar o conceito de *modelos lineares generalizados*, que consiste em três componentes:

- Um componente aleatório, que contém a distribuição de probabilidade da variável dependente (Y).
- Uma componente sistemática, a qual corresponde a uma função linear entre as variáveis independentes.
- Uma função de ligação, que é responsável por descrever a relação matemática entre a componente sistemática e a componente aleatória. Existem duas classes de função de ligação, denominadas *logit* e *log-linear* [Dobson 1990].

O modelo de regressão logística binária é um caso especial de modelo linear generalizado, que usa a função de ligação logit para obter as estimativas dos coeficientes da Equação 1 [Venables et al. 2009].

Assim é possível obter um modelo de regressão logística. Entretanto, se torna necessário verificar quais variáveis são mais significativas para o modelo, já que modelos com muitos parâmetros apresentam correlações entre as variáveis e grande variação nas estimativas dos coeficientes. Para reduzir o modelo nós usamos a técnica *Stepwise*, a qual permite selecionar e remover variáveis, que são menos significantes para o modelo [McCullagh and Nelder 1989]. Finalmente é possível encontrar a probabilidade de sucesso, aplicando na Equação 1 os valores das estimativas dos coeficientes encontrados.

4. Metodologia

O uso de regressão logística para criar um *ranking* de credibilidade, iniciou-se com o tratamento e a adequação da base de dados, a qual será descrita na seção 5.1. Nessa base de dados, cada transação tem sua respectiva variável resposta e outros atributos independentes. A fim de facilitar a análise dos dados todos os atributos foram normalizados.

Para construir o modelo de regressão utilizamos a ferramenta de software R [Team 2009], que é um software livre que possui diversos pacotes estatísticos. Para aplicar a regressão logística usamos o pacote *GLM* (*generalized linear models*), contido no software. Foi definido um atributo como variável resposta e os demais como variáveis independentes. É importante explicar que essa variável resposta é o *feedback* da transação. Para determinar que será usada uma regressão logística binária, foram configurados os campos FAMILY como binomial e o LINK como *logit*.

A fim de encontrar o melhor modelo, foram removidas as variáveis independentes menos significativas para o modelo. Sendo que o melhor modelo é o que apresenta o menor índice de Akaike (AIC). Usamos a técnica de *Stepwise* para realizar esta otimização. Após definir os melhores modelos foi possível obter as estimativas dos coeficientes para os atributos independentes. Em seguida foi calculado a probabilidade de uma transação obter *feedback* positivo(1), usando a Equação 1.

A metodologia usada em nosso trabalho pode ser melhor entendida pelo diagrama descrito na Figura 1. Para a construção do *ranking* de credibilidade, ordenamos em ordem decrescente as probabilidades calculadas pelo modelo de regressão.

A fim de verificar a qualidade dos modelos, escolhemos 1000 registros nos extremos (fundo e topo) do *ranking*, já que essas são as partes mais relevantes para o modelo (onde gostaríamos que estivesse localizado a maior e menor parte de transação confiáveis, respectivamente).

Cada modelo de credibilidade produz um *ranking*, onde cada posição nesse *ranking* é uma transação que contém uma probabilidade de obter *feedback* positivo e o *feedback* real obtido do usuário, que representa a variável resposta.

Espera-se que o maior número de *feedback* positivo esteja no topo do *ranking*. Analogamente é esperado que um menor número de *feedback* positivo se encontre no fundo do *ranking*. Assim, poderemos obter a precisão de cada modelo, avaliando diferentes faixas do *ranking*, comparando o valor estimado com o valor real.

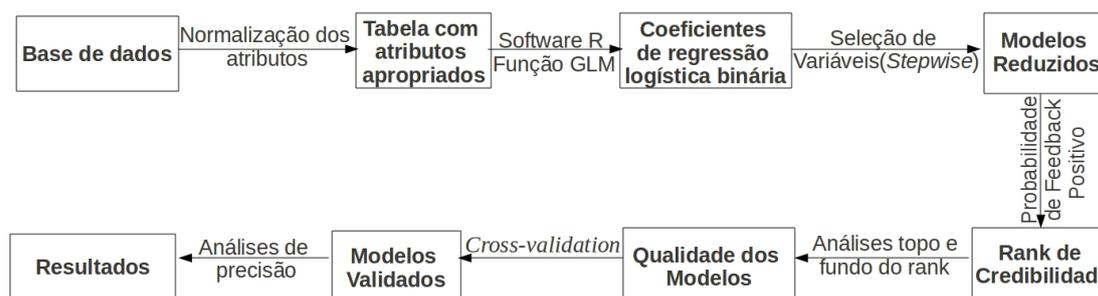


Figura 1. Modelagem da Credibilidade - Definição do Processo

A técnica de *K-fold cross-validation* foi usada para testar a qualidade de cada modelo. Definimos como 5 o número de sub-amostras (K). Assim, o conjunto de dados foi dividido em 5 partes uniformes, onde cada uma foi usada como dados de validação para encontrar os coeficientes do modelo. As outras 4 sub-amostras foram usadas como dados de treinamento, onde o modelo foi aplicado. A precisão foi calculada em cada uma das sub-amostras de treinamento, seguindo os mesmos métodos explicados para a base de dados completa. A precisão final foi obtida através da média aritmética de cada conjunto de valores.

5. Estudo de Caso

Esta seção apresenta nosso estudo de caso, onde aplicamos nossa metodologia em uma base de dados real de um mercado eletrônico para avaliar alguns modelos de credibilidade. Inicialmente, descreveremos brevemente o conjunto de dados na seção 5.1, apresentando em seguida os resultados na seção 5.2.

5.1. Visão geral da base de dados

O *TodaOferta*¹ [Pereira et al. 2009] é um mercado eletrônico desenvolvido, para compra e venda de serviços e produtos através da web, pelo maior provedor de serviços de Internet na América Latina, o Universo Online Inc. (UOL)².

A Tabela 1 apresenta um pequeno resumo da base de dados do *TodaOferta*. Essa amostra possui uma quantidade significativa de usuários, anúncios e negociações. Devido ao acordo de confidencialidade dos dados, as informações quantitativas não serão apresentadas neste artigo. O subconjunto da base de dados que iremos utilizar nessa pesquisa consiste de algumas dezenas de milhares de transações.

Cobertura (tempo)	Jul/2007 to Jul/2009
#categorias (<i>top-level</i>)	32
#sub-categorias	2189
Média de anúncios por vendedor	42,48
Opções de negociação	Preço fixo e Leilão

Tabela 1. TodaOferta - Sumário da Base de Dados

O mercado eletrônico *TodaOferta* utiliza um simples mecanismo de reputação. Após cada negociação, compradores e vendedores qualificam a outra parte de forma positiva (1), negativa (-1) ou neutra (0). A reputação dos usuários é definida como a soma

¹<http://www.todaoferta.com.br>

²<http://www.uol.com.br>

de todas as qualificações recebidas por ele. Os *feedbacks* de um mesmo usuário só são consideradas uma única vez para computar o *score* da reputação.

Ofertas são criadas por vendedores para anunciar produtos ou serviços. As ofertas podem ser ofertadas com um preço fixo ou como um leilão. Ao contrário do *eBay*, onde leilões representam cerca de 50% do total de transações [Holahan 2008], no *TodaOferta* leilões representam menos de 2% do total de transações concretizadas.

Existem 32 níveis de categorias no toda oferta, o qual incluem 2189 sub-categorias provendo uma variedade de produtos e serviços distintos. Atualmente o maior número de vendas são de telefone celulares, *MP3 players* e *pen-drives*.

Selecionamos 15 atributos da base de dados para serem usados como candidatos para o modelo de regressão logística, que estão descritos a seguir.

- **Preço:** Preço do produto/serviço sendo oferecido..
- **Duração:** Duração da oferta (anúncio) estabelecido pelo vendedor (em dias).
- **Destaque:** Indica se a oferta foi anunciada com destaque (pacote de anúncios especiais).
- **Visualizações:** O número de visualizações da oferta.
- **Oferta com Pagamento Seguro:** Indica se uma oferta tem a opção de utilizar o mecanismo de pagamento seguro provido pelo mercado eletrônico.
- **Transação segura:** Identifica a transação que foi efetuada utilizando o mecanismo de pagamento seguro.
- **Itens Vendidos:** O montante total de itens que o vendedor vendeu no mercado eletrônico.
- **Tempo de Cadastro:** Há quanto tempo o vendedor está registrado no mercado eletrônico.
- **Qualificações Positivas:** O montante de qualificações positivas um usuário(vendedor) recebeu.
- **Porcentagem de Qualificações Positiva :** O montante relativo de qualificações positivas que um usuário(vendedor) recebeu.
- **Qualificação Global:** A soma das qualificações do vendedor, considerando diferentes tipos de qualificações recebidas.
- **Valor Total Negociado:** O montante total de dinheiro negociado pelo vendedor no mercado eletrônico.
- **Valor médio negociado:** A média do preço por transação executada pelo vendedor.
- **Mega Vendedor:** Indica se o usuário é considerado um vendedor de destaque pelo *TodaOferta*.
- **Selo Qualidade:** Denota o vendedor que possui um certificado de qualidade, provido por uma empresa à parte.

5.2. Resultados

A otimização usando a técnica de *stepwise* para construir o melhor modelo de regressão logística resultou em diferentes modelos, alguns desses modelos se adaptaram melhor para o topo do *ranking*, outros para o fundo. A seguir iremos apresentar os modelos, onde cada modelo é composto pelos atributos que melhor influenciaram no valor do *feedback*.

O modelo *Baseline* (linha de base) é formado pelos atributos *Porcentagem de Qualificações positivas* e *Qualificação Global*, os quais são considerados os atributos básicos

para gerar o feedback do mercado eletrônico usado no nosso estudo de caso. A fim de melhorar esse modelo, adicionamos novos atributos, preocupando sempre em preservar o valor do *Akaike*. Assim, 4 novos modelos foram criados.

O **modelo A** é composto pelas variáveis do *Baseline* e o atributo *Destaque*. O **modelo B** foi gerado pelo acréscimo da variável *Mega Vendedor* ao *Baseline*. O **modelo C** é composto pelos atributos do *modelo A* com o acréscimo da variável *Visualizações*. Já o **modelo D** foi criado a partir dos atributos do *modelo B* com incremento das variáveis *Itens Vendidos*, *Tempo de Cadastro* e *Oferta com Pagamento Seguro*. Além dos modelos gerados por regressão logística, foi criado um modelo aleatório para facilitar a comparação e análise com os outros modelos, já que o **modelo Aleatório** expressa a probabilidade geral de acordo com a base de dados utilizada.

Analisando os modelos, foi possível calcular a precisão em diferentes partes do *ranking*, onde nos focamos a análise sobre o topo e o fundo do *ranking*, como havíamos explicado. O modelo *Baseline* foi mais preciso que o *modelo Aleatório* em todos os intervalos avaliados. Entretanto o modelo *Baseline* apresentou menores valores de precisão em posições intermediárias do *ranking*.

Os modelos *A* e *C* não foram precisos ao prever valores de probabilidades para o topo do *ranking*. Contudo obtiveram uma boa precisão no extremo fim (fundo) do *ranking*, superando o modelo *Baseline* em muitas faixas do *ranking*.

Os modelos *B* e *D* apresentaram melhor precisão quando usados para prever valores de probabilidades para os integrantes do topo do *ranking*. Esses modelos foram melhores que o *Baseline* em maior parte das faixas do *ranking*.

A fim de avaliar os modelos, comparamos a precisão dos *rankings* obtidos para topo e fundo (Figuras 2 and 3). Esses resultados foram construídos a partir da análise dos dados gerados pela técnica de *K-fold-Cross Validation*. Foram analisados 1000 registros nos extremos (fundo e topo) e gerado um gráfico para cada extremo. Esses comparam os modelos criados pela regressão logística com o *modelo Aleatório*, em faixas variando de 10 em 10 registros (ver gráficos 3 e 2).

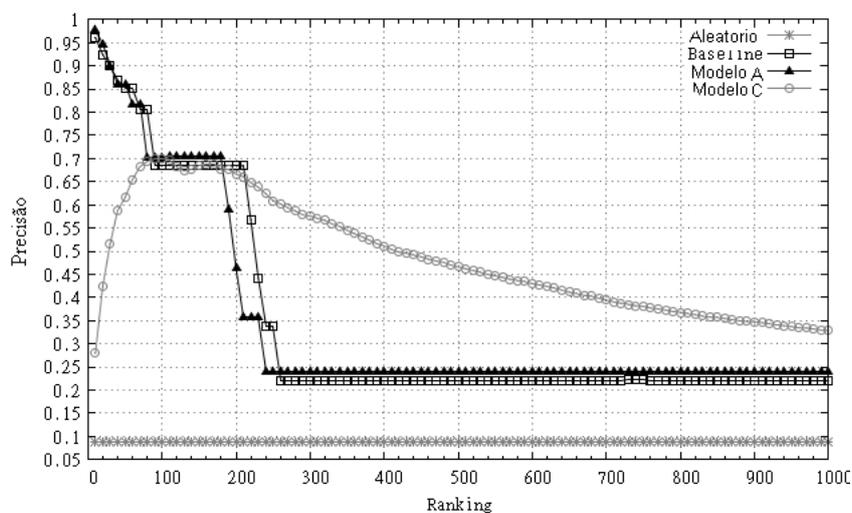


Figura 2. Modelos de credibilidade - Precisão no fundo do *ranking*.

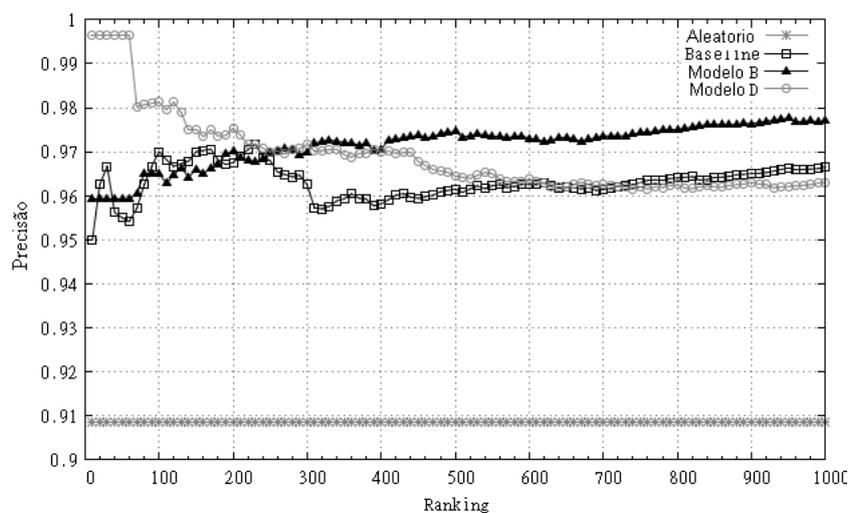


Figura 3. Modelos de Credibilidade - Precisão no topo do ranking.

Analisando o gráfico de precisão para o fundo (Figura 2), podemos observar que o *modelo A* obteve melhor precisão no extremo fim (correspondente aos 180 últimos registros) do *ranking*. A melhoria máxima obtida pelo modelo, em comparação ao *Baseline*, foi de 2,3% de precisão, sobre um máximo de 7,6%. O *modelo C* obteve melhor precisão depois da posição dos 220 registros do *ranking*. A melhoria máxima obtida pelo modelo, em comparação ao *Baseline*, foi de 38,2% , sob um máximo de 78%. Podemos observar que o modelo *Baseline* foi o mais preciso entre as faixas de 180 a 220 do *ranking*.

No gráfico de precisão para o Topo (Figura 3), podemos notar que o *modelo D* foi mais eficaz que o *Baseline* nas 600 primeiras posições do *ranking*, ou seja, maiores probabilidades de obter *Feedback* positivo. O *modelo D* alcançou 99,6% de precisão nos 60 primeiros registros e apresentou uma melhoria máxima de 4,6%, em comparação ao *Baseline*, sob um máximo de 5%. O *modelo B* começou a se destacar a partir da posição 250. No intervalo de 250 a 1000, o modelo foi melhor do que os modelos *D* e o *Baseline*. A melhoria máxima em relação ao *Baseline* foi de 1,5% (ante um teto máximo de 3%). Portanto, ao analisar o topo do *ranking*, os novos modelos de credibilidade superaram o modelo *Baseline* em todas as faixas do *ranking*, indicando um melhor probabilidade de obter *feedback* positivo para transações no topo desses modelos de *ranking*.

6. Conclusão

A popularização da Web tem oferecido novos serviços a cada dia, demandando mecanismos para assegurar a credibilidade desses serviços. Até o momento, pouco tem sido feito para medir e entender a credibilidade desse complexo ambiente Web, o que por si só é um grande desafio de pesquisa.

Mercados eletrônicos constituem importante cenário de pesquisa devido à sua popularidade e volume de negócios nos últimos anos. Nesse cenário a reputação assume um importante papel, principalmente para proteger compradores de vendedores fraudulentos. Um mecanismo de reputação procura provê um indicador de quão confiável é o usuário, com base em seu histórico de transações realizadas.

Neste trabalho, usamos regressão logística para modelar e avaliar a credibilidade de aplicações Web. Essa avaliação é baseada em uma amostra representativa de serviços e suas respectivas avaliações dos usuários e um *ranking* que representa uma escala de credibilidade gerada pelo modelo. Chamamos de modelo de credibilidade uma função capaz de determinar o valor de credibilidade para uma transação de uma aplicação Web, considerando diferentes critérios desses serviços e o usuário fornecedor do serviço. Quanto maior a capacidade dos modelos em posicionar vendedores que oferecem serviços satisfatórios (que são bem qualificados a partir do *feedback*) no topo desse *ranking*, melhor é a qualidade do modelo.

Executamos experimentos usando uma base de dados real de um mercado eletrônico, a partir do qual avaliamos o modelo de regressão logística usando diferentes tipos de fonte de informação, tais como atributos relacionados às características das ofertas, *expertise* do vendedor e sua qualificação. Os resultados mostram que nosso trabalho pode ser muito proveitoso e promissor. Os resultados obtidos foram muito bons, apresentando ganhos representativos, quando comparados ao *baseline*. Observamos que existem diferentes modelos para o topo e fundo do *ranking*, assim nos executamos uma análise diferente a fim de classificar as transações online em ambos os cenários.

Como trabalhos futuros queremos melhorar a avaliação e análises dos modelos de credibilidade apresentados neste trabalho. Pretendemos implementar novos modelos de credibilidade baseado em técnicas de aprendizado de máquina e algoritmos genéticos.

Agradecimentos

Este trabalho foi parcialmente patrocinado pelo Universo OnLine S. A. - UOL (www.uol.com.br) e parcialmente patrocinado pelo Instituto Nacional de Ciência e tecnologia para a Web (CNPq no. 573871/2008-6), CAPES, CNPq, Finep, e Fapemig.

Referências

- Agresti, A. (1996). *An Introduction to Categorical data Analysis*. John Wiley and Sons, New York.
- Amin, A., Zhang, J., Cramer, H., Hardman, L., and Evers, V. (2009). The effects of source credibility ratings in a cultural heritage information aggregator. In *WICOW '09: Proc. of the 3rd workshop on Information credibility on the web*, pages 35–42, NY, USA. ACM.
- Casella, G. and Berger, R. (2002). *Statistical Inference*. Pacific Grove, 2nd edition.
- Dellarocas, C. (2006). Reputation mechanisms. In *Handbook on Economics and Information Systems*, pages 629–660. Elsevier Publishing.
- Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. London:Chapman and Hall.
- Flanagin, A. J. and Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society*, 9(2):319–342.
- Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In *WWW '04: Proc. of the 13th international conference on World Wide Web*, pages 403–412, NY, USA. ACM.

- Holahan, C. (2008). Auctions on ebay: A dying breed. *BusinessWeek online*.
- Hosmer, D. W. (2000). *Applied Logistic Regression*. Wiley, New York, 2nd edition.
- Houser, D. and Wooders, J. (2006). Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369.
- Jøsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decis. Support Syst.*, 43(2):618–644.
- Juffinger, A., Granitzer, M., and Lex, E. (2009). Blog credibility ranking by exploiting verified content. In *Proc. of the 3rd workshop on Information credibility on the web*, pages 51–58, NY, USA. ACM.
- Klos, T. and Alkemade, F. (2005). Trusted intermediating agents in electronic trade networks. In *AAMAS '05: Proc. of the 4th international joint conference on Autonomous agents and multiagent systems*, pages 1249–1250, New York, NY, USA. ACM.
- Lima, R. and Pereira, A. (2011). Applying logistic regression to rank credibility in web applications. In *7th International Conference on Web Information Systems and Technologies*.
- Mccullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, 2nd edition.
- Melnik, M. I. and Alm, J. (2002). Does a seller's ecommerce reputation matter? evidence from ebay auctions. *Journal of Industrial Economics*, 50(3):337–49.
- Pereira, A. M., Duarte, D., Jr., W. M., Almeida, V., and Góes, P. (2009). Analyzing seller practices in a brazilian marketplace. In *18th International World Wide Web Conference*, pages 1031–1041.
- Resnick, P. and Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce*, Elsevier Science B.V.:127–157.
- Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60.
- Team, R. (2009). R: A language and environment for statistical computing. <http://www.r-project.org>. The R Development Core Team.
- Venables, W. N., Smith, D. M., and the R Development Core Team (2009). An introduction to r. <http://www.cran.r-project.org>.