APPLYING LOGISTIC REGRESSION TO RANK CREDIBILITY IN WEB APPLICATIONS

Rafael Lima and Adriano Pereira
Dept. of Computer Engineering (DECOM), Federal Center for Technological Education of Minas Gerais (CEFET-MG)
Belo Horizonte - MG, Brazil

Keywords: Web applications, Credibility, Ranking, Trust, e-Markets, Web 2.0.

Abstract: The popularization of the World Wide Web (WWW) has given rise to new services every day, demanding mechanisms to ensure the credibility of these online services. Since now, little has been done to measure and understand the credibility of this complex Web environment, which itself is a major research challenge. In this work, we use logistic regression to design and evaluate the credibility of a Web application. We call a credibility model a function capable of assigning a credibility value to transaction of a Web application, considering different criteria of this service and its supplier. 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, and it allows to compare and evaluate these credibility models. 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.

1 INTRODUCTION

The popularization of Web 2.0 applications, where users can interact more, creating and sharing a diversity of content, trading products and establishing new communities, represents a major revolution in how users and corporations use the Web. This revolution has brought challenges related to credibility, pertaining to the usage of these Web applications or services. Thus, mechanisms that help users to evaluate credibility, when using these services, has become essential.

Digital libraries, e-markets, user-generated content and sharing systems are examples of Web applications that require mechanisms for assessing credibility. Many of these applications already provide systems to deal with this, such as reputation systems.

Evaluating and quantifying credibility in a Web application represents the major challenge of this research. Among the main difficulties of this task, we can highlight the large number of variables involved and the low reliability of the information available.

Models of credibility differ from reputation models, which are widely studied in the literature (Jøsang et al., 2007; Sabater and Sierra, 2005),because they not only consider feedback from users, but also a set of attributes, which can be related to the service provided and its supplier, as a way to get a more complete and effective evaluation of a given service available on the Web.

It is important to explain that, despite the problems of reputation systems (Resnick et al., 2000), it is necessary to use feedback information to measure the user opinion of a service that can be described by different characteristics that we denote credibility attributes in our model. Moreover, there are specific works that deal with improving the quality of reputation systems, such as the identification of fraudsters of these systems (Maranzato et al., 2010), which was also used in the real application used in this research.

In this work, we use logistic regression to design and evaluate the credibility of a Web application. This evaluation is based on a representative sample of services that have user feedbacks and a ranking that represent a scale of credibility generated by the model. The greater the capacity of the model to position vendors that offer satisfactory services (which are qualified as such from the feedbacks) in the top positions on this scale, the higher its quality. We perform experiments using an actual dataset of an electronic market, from which we evaluate the logistic regression model using different types of information sources, such as attributes related to offer's characteristics, seller's expertise and qualification. The results show that our approach can be very useful and promising. The ob-

Lima R. and Pereira A..
APPLYING LOGISTIC REGRESSION TO RANK CREDIBILITY IN WEB APPLICATIONS.
DOI: 10.5220/0003345104800485
In *Proceedings of the 7th International Conference on Web Information Systems and Technologies* (WEBIST-2011), pages 480-485
ISBN: 978-989-8425-51-5

tained results were very good, showing representative gains, when compared to a baseline.

2 RELATED WORK

In the recent years, the concept of credibility has begun to be studied on the Web, in order to measure whether a user relies on a service or information available. It is a consensus in the literature that credibility can be subjective to the user, but it also depends on objective measures. The credibility of Web applications has become a multidisciplinary subject, where researchers from communication have been focusing on a more qualitative (and subjective) analysis of credibility (Flanagin and Metzger, 2007), while the area of computer science has focused on more objective metrics. The methods proposed in the area of computer science are strongly based on trust and reputation (Guha et al., 2004), and credibility rankings that take into account the source of information (Amin et al., 2009) and its content (Juffinger et al., 2009).

Reputation mechanisms are based on virtual opinions, given by people who generally do not know each other personally. Therefore, electronic trust is more difficult to be established if compared to real world trust. Taking a broad view, in these marketplaces a buyer's reputation represents the probability of payment and a seller's reputation represents the probability of delivering the advertised item (product that has been bought) after the payment (Houser and Wooders, 2006). These probabilities are related to trust (Melnik and Alm, 2002).

Electronic markets are getting more popular each day. Several works investigate reputation systems and how they induce cooperative behavior in strategic settings. Dellarocas (Dellarocas, 2006) has done a thorough review on this topic. While providing incentive to good behavior, reputation systems may also help eliciting deceptive behavior.

Klos et. al (Klos and Alkemade, 2005) analyze the effect of trust and reputation over the profits obtained by intermediaries in electronic commercial connections. Different trust and distrust propagation schemes in e-commerce negotiations are studied and evaluated in Guha et. al (Guha et al., 2004).

Resnick et al. (Resnick et al., 2000) say that these reputation systems have three main problems: (i) buyers have little motivation to provide feedback to sellers; (ii) it is difficult to elicit negative feedback because it is common that people negotiate and solve problems before filling the evaluation in the system; (iii) it is difficult to assure honest reports. Since it is very simple to register in such systems, it is very easy to create a false identity that can be used to trade with other users and distort the reputation system.

The researches that we describe in this section suggests the increasing need of providing new credibility models that provide subsidies to users of online services in order to allow them to act with more confidence an trust in the Web.

3 LOGISTIC REGRESSION

Logistic regression is a statistical technique that produces from a set explanatory variables, a model that can predict values taken by a categorical dependent variable. Thus, a regression model is able to calculate the probability of an event, through the *link* function described by the following Equation:

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

where $\pi(x)$ is the probability of success when the value of the predictive variable is x. β_0 is a constant used for adjustment and β_i are the coefficients of the predictive variables (Hosmer, 2000). To find the estimation of coefficients *beta* in Equation 1, the maximum likelihood technique is used. This maximizes the probability of obtaining the data group observed through estimated model. In logistic regression this technique can be resolved by *Newton-Raphson* method (Casella and Berger, 2002).

The regression model can be of ordinal or nominal nature,depending on the values that the dependent variable can assume (Agresti, 1996). In this project, the nominal logistic regression will be used because there is no order between the categories of the variable. In this project, the dependent variable contains two categories (Dichotomous variable). Therefore, we used a binary regression logistic model with multivariable,i.e., more than one independent variable.

In order perform the logistic regression,it is important to explain the concept of *generalized linear models* (GLM). This consists of three components:

- A random component, which contains the probability distribution of the dependent variable (Y).
- A systematic component, which corresponds to a linear function between the independent variables.
- A *link* function, that is responsible for describing the mathematical relationship between the systematic component and random component.

There are two classes of *link function*, log-linear and logit. In logistic regression, the function logit is used. (Dobson, 1990) .

The binary logistic regression model is a special case of the GLM model with the logit function. This function is used to get the estimation of coefficients of the Equation 1 (Venables et al., 2009). Thus, is possible to obtain a logistic regression model. Moreover, it is necessary to check which variables are most significant for the model, since models with many variables show a correlation between the variables and large variation in estimation of the parameters.

We use *stepwise* technique to reduce the model, which allows the selection and removal of variables, that are less significant for the model (Mccullagh and Nelder, 1989).Finally it is possible to find the probability of success,using the values of estimated coefficients in Equation1.

4 METHODOLOGY

The use of logistic regression to create a credibility rank initiate with the pre-process of the dataset, which will be described in Section 5.1. In this dataset, each transaction has its respective response variable and other independent attributes. The attributes are normalized in order to make easier the data analysis.

We use *R* software tool (Version, 2009) to build the logistic regression model, which is a free software that has several statistical packages. In order to apply the binary logistic regression the *GLM* (generalized linear models) package was used. One attribute was defined as the response variable and the other attributes as independent variables. It is important to explain that this response variable is the *feedback* of a transaction. The configuration field FAMILY was set as binomial and the LINK as logit.

In order to find the best model, the less significant independent variables were removed. The best model has the lowest *Akaike*(AIC). We use the *Stepwise* technique to perform this optimization. After defining the best model, it was possible to get the estimation of coefficients for the independent attributes. This way, the odds (chance) of a transaction to achieve positive feedback is calculated using Equation 1.

The methodology used in our work can be better understood by the workflow described in Figure 1.

To build the credibility rank, the odds calculated by models were sorted in descending order. Thus, at the bottom of the ranking are the smallest odds of getting positive feedback. By contrast, the records with highest odds are located on the top of ranking.

In order to verify the quality of the prediction's models, we choose the 1,000 records from the top and the bottom of the ranking, since these are the most relevant parts of it (where would be located the most and the least trustable transactions, respectively).

Each credibility model will produce a ranking, where each position of the ranking is a transaction that has a probability of getting a positive feedback and

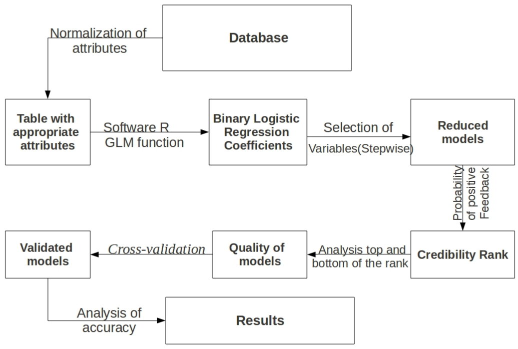

Figure 1: Credibility Rank- Definition process.

the real *feedback* that represents the response variable. It is expected that the highest probability values are located at the top of the ranking. Moreover, analogously, at the bottom of the ranking, it is expected to find the smallest probability values. Thus, we can obtain the precision of each model evaluating different ranges of the ranking, comparing the estimated and actual values.

The technique of *K-fold cross-validation* was used for testing the quality of each credibility model. We define as five the number of subsamples (*K*). Thus, the dataset was divided in 5 uniform parts, where each part was used as a validation data, that is, to find the coefficients of the model. The other four sub-samples were used as training data, where the model was applied. The precision is calculated in each of the subsamples following the same method explained for the whole dataset. The final value is calculated through an arithmetic mean of each set of values.

5 CASE STUDY

This section presents our case study where we apply our methodology to evaluate some credibility models using actual data from an electronic market. First we briefly describe the dataset in Section 5.1, presenting the results in Section 5.2.

5.1 Dataset Overview

TodaOferta[1] (Pereira et al., 2009), which is a marketplace developed by the largest Latin America Internet Service Provider, named Universo Online Inc. (UOL)[2], is a website for buying and selling products and services through the Web.

Table 1 shows a short summary of the *TodaOferta* dataset. It embeds a significant sample of users, list-

[1]http://www.todaoferta.com.br
[2]http://www.uol.com.br

ings, and negotiations. Due to a confidentiality agreement, the quantitative information about this dataset can not be presented. The subset of this dataset that we have used in this research comprises some tens of thousands of transactions.

Table 1: TodaOferta Dataset - Summary.

Coverage (time)	Jul/2007 to Jul/2009
#categories (top-level)	32
#sub-categories	2,189
Average listings per seller	42.48
Negotiation options	Fixed Price and Auction

In *TodaOferta*, buyers are users, listings are services, and sellers are service providers. The *TodaOferta* marketplace employs a quite simple reputation mechanism. After each negotiation, buyers and sellers qualify each other with a rate of value 1 (positive), 0 (neutral), or -1 (negative). User's reputation is defined as the sum of all qualifications received by him/her. Feedbacks from a same user are considered only once when computing the reputation score. Reputation systems are useful to communicate trust in electronic commerce applications. However, *TodaOferta* provides other information about sellers and buyers that can be as well used to identify trustful and distrustful users (e.g., time since the user is registered, comments left by users who negotiated with him/her).

Listings are created by sellers to advertise products or services. Listings can be offered at a fixed-price or as an auction. When a buyer is interested in a listing he/she starts a negotiation. In the case of a fixed-price listing, the negotiation automatically generates a transaction, meaning that buyer and seller should transact the good at the advertised price. If the listing is an auction, the winning bid will become a transaction when the auction finishes. Unlike eBay, where auctions generate almost 50% of all transactions (Holahan, 2008), in *TodaOferta* auctions represent less than 2% of all transactions, since the vast majority of listings are fixed-price.

There are 32 top-level categories in *TodaOferta*, which include 2,189 sub-categories providing a variety of distinct products and services, from collectibles to electronics. The current top sales sub-categories are cell phones, MP3 players and pen drives.

From the this dataset we select 15 attributes to be used as candidates for the logistic regression model:

- **Price**: price of the product/service being offered.
- **Duration**: duration of the listing (product ad) set by the seller(in days).
- **Highlight**: indicates whether the listing is set to be advertised with highlight (some special advertisement package).
- **Views**: the number of visualizations of the listing.
- **Offer with SafePayment**: indicates whether a listing or offer has the option of using a safe payment mechanism provided by the e-market.
- **Safe Transaction**: identifies a transaction that is performed adopting the safe payment mechanism.
- **Sold Items**: the amount of items the seller has already sold in the e-market.
- **Registration Time**: how long the seller has been registered in the e-market.
- **Positive Qualifications**: the amount of positive qualifications a user (seller) has received.
- **Percentage Positive Qualifications**: the relative amount of positive qualifications a user (seller) has received.
- **Global Score**: the seller reputation rating score, considering the different score types.
- **Total Negotiated Value**: the total amount of money negotiated by the seller in the e-market.
- **Average Negotiated Value**: the average price per transaction performed by the seller.
- **Retailer**: indicates whether the user is considered a powerful seller by *TodaOferta*.
- **Certified**: denotes the seller who has a certification of quality, which is provided by a third party company.

5.2 Results

The optimization using *Stepwise* technique to build a best logistic regression model results in different models, some of them more suitable for the top of the ranking and other ones for the bottom of the ranking. Below we present the models, where each model is composed by attributes that showed greatest influence on the value of feedback.

The **Baseline** model is formed by the attributes *Percentage Positive Qualifications* and *Global Score*, which are considered the most significant variables to generate the basic *Feedback* of the e-market we used as case study. To improve this model, new attributes were added, preserving and improving the value of *Akaike*. Thus, four new models were built.

Model A consists of variables of the Baseline model and the attribute *Highlight*. **Model B** was generated by adding variable *Retailer* to Baseline model. **Model C** consists of attributes of the model A, adding the variable *Views*. **Model D** was created from the attributes of model B with addiction of variables *Sold Items*, *Registration Time* and *Offer with SafePayment*.

Besides the logistic regression models, a random model was created to make easier the comparison and analysis with other models. The random model expresses the probability of finding a record with suitable *feedback*(positive or negative) to scale of the ranking (top or bottom), regardless of the variables used. In other words, this model indicates the percentages of positive and negative feedback that are observed in the dataset.

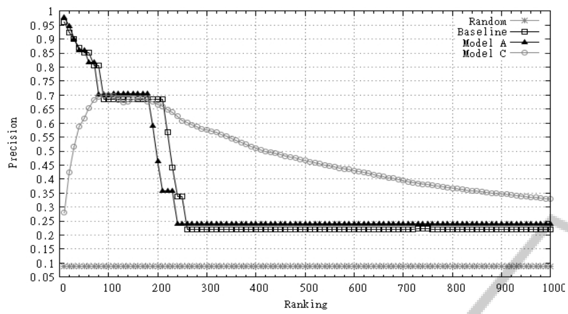

Figure 2: Credibility models - Precision at bottom.

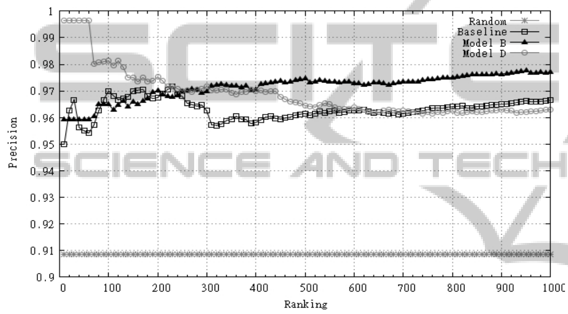

Figure 3: Credibility models - Precision at top.

Evaluating the models, it was possible to calculate the precision at different parts of the ranking, where we focus on the top and bottom, as we have already explained. The Baseline model was more accurate than the random model in all evaluated intervals of the ranking. However, the Baseline model presented smaller precision values in intermediate positions of the ranking. A similar behavior was observed for the bottom of it.

Models A and C were not accurate in predicting values of probability for the top of the ranking. However, they obtained a good precision at the extreme end (bottom) of the ranking, surpassing the Baseline model in most of the scales of the rank.

Models B and D have presented higher accuracy in predicting values of probability for the top of the ranking positions. These models were better that the Baseline in most parts of the ranking.

In order to evaluate the models we create two graphs (Figures 2 and 3) of precision x ranking, each one with focus on these 1,000 top or bottom positions of the ranking. These results were built from data analysis generated by the K-fold-Cross Validation technique. The graphs compare the accuracy of the models created using logistic regression with the random model.

Analyzing the graph of precision at the bottom of the ranking (Figure 2), we can observe that the model A is the best one for the end of the ranking (that is, the last 180 records). The maximum improvement obtained by the model, in comparison with the baseline, was 2.3% of accuracy, under a maximum of 7.6%. The C model is the best one after the position 220 of records of the ranking. The maximum improvement obtained by this model, in comparison to the baseline, was 38.2% of accuracy, under a maximum of 78.0%. We can observe that the baseline model was the best one in the range from 180 to 220 of the ranking.

In the graph of precision for top (Figure 3), we can see that model D was more effective than the baseline model in the 600 first positions of the ranking, showing more probability to get positive *Feedback*. This model achieved 99.6% of accuracy for the 60 first records and showed a maximum improvement of 4.6%, in comparison to baseline, under a maximum of 5%. The model B was the best one after the 250 initial positions of the ranking. In these interval from 250 to 1,000, it was more accurate than D and Baseline models. The maximum improvement obtained by this model was 1.5% of accuracy, under a maximum of 4.3%. Therefore, considering the top of the ranking, our new credibility models overcome the accuracy of baseline model in all ranges of the ranking, indicating a higher probability of positive feedback for transactions at the top of these ranking models.

The next section presents the conclusions of our work and future directions for this research.

6 CONCLUSIONS

The popularization of Web has given rise to new services every day, demanding mechanisms to ensure the credibility of these services. Since now, little has been done to measure and understand the credibility of this complex Web environment, which itself is a major research challenge.

E-markets constitute an important research scenario due to their popularity and revenues over the last years. In this scenario, reputation plays an important role, mainly for protecting buyers from fraudulent sellers. A reputation mechanism tries to provide an indication of how trustworthy a user is, based on his/her performance in previous transactions.

In this work, we use logistic regression to design and evaluate the credibility of a Web application. This evaluation is based on a representative sample of services that have user feedbacks and a ranking that represent a scale of credibility generated by the model. We call a credibility model a function capable of assigning a credibility value to transaction of a Web application, considering different criteria of this service and its supplier. The greater the capacity of the model to position vendors that offer satisfactory services (which are qualified as such from the feedbacks) in the top positions on this scale, the higher its quality.

We perform experiments using an actual dataset of an electronic market, from which we evaluate the logistic regression model using different types of information sources, such as attributes related to offer's characteristics, seller's expertise and qualification. The results show that our approach can be very useful and promising. The obtained results were very good, showing representative gains, when compared to a baseline. We observe that there are different models for the top and the bottom of the ranking, thus we perform a different analysis in order to identify the best solutions obtained to rank the online transactions in these both scenarios.

These results motivate further work, showing there are much more to analyze and conclude about these credibility models and how to combine even better these models to generate other ones that can be more reliable and that can help users to perform safe transactions on the Web.

As future work we want to improve the evaluation and analysis of the credibility models that we have presented in this work. Moreover, we want to implement new credibility models based on techniques of machine learning and genetic algorithms.

ACKNOWLEDGEMENTS

This work was partially sponsored by Universo OnLine S. A. - UOL (www.uol.com.br) and partially supported by the Brazilian National Institute of Science and Technology for the Web (CNPq grant no. 573871/2008-6), CAPES, CNPq, Finep, and Fapemig.

REFERENCES

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:Duxbury, 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. B. and Alkemade, F. (2005). Trusted intermediating agents in electronic trade networks. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1249–1250, New York, NY, USA. ACM.

Maranzato, R., Pereira, A., do Lago, A. P., and Neubert, M. (2010). Fraud detection in reputation systems in e-markets using logistic regression. In *SAC '10: Proc. of the 2010 ACM Symposium on Applied Computing*, pages 1454–1459, New York, NY, USA. ACM.

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., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Commun. ACM*, 43(12):45–48.

Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60.

Venables, W. N., Smith, D. M., and the R Development Core Team (2009). An introduction to r. http://www.cran.r-project.org.

Version, T. R. D. C. T. (2009). R: A language and environment for statistical computing. http://www.r-project.org.