Personalizing Trust in Online Auctions

John O'Donovan [a,1], Vesile Evrim [b] Barry Smyth [a] Dennis McLeod [b] Paddy Nixon [a]
[a] *School of Computer Science and Informatics, University College Dublin, Ireland.*
[b] *Semantic Information Research Laboratory. Viterbi School of Engineering, University of Southern California, Los Angeles.*

Abstract. The amount of business taking place in online marketplaces such as eBay is growing rapidly. At the end of 2005 eBay Inc. reported annual growth rates of 42.5% [3] and in February 2006 received 3 million user feedback comments per day [1]. Now we are faced with the task of using the limited information provided on auction sites to transact with complete strangers with whom we will most likely only interact with once. People will naturally be comfortable with old fashioned "corner store" business practice [14], based on a person to person trust which is lacking in large-scale electronic marketplaces such as eBay and Amazon.com. We analyse reasons why the current feedback scores on eBay and most other online auctions are too positive. We introduce *AuctionRules*, a trust-mining algorithm which captures subtle indications of negativity from user comments in cases where users have rated a sale as positive but still voiced some grievance in their feedback. We explain how these new trust values can be propagated using a graph-representation of the eBay marketplace to provide personalized trust values for both parties in a potential transaction. Our experimental results show that *AuctionRules* beats seven benchmark algorithms by up to 21%, achieving up to 97.5% accuracy , with a false negative rate of 0% in comment classification tests compared with up to 8.5% from other algorithms tested.

Keywords. Trust, Transitivity, Online Auctions, Personalization

1. Introduction

The majority of feedback comments on online auction transactions are positive. [14]. According to our analysis eBay is over 99% biased towards positive comments. Currently, eBay compiles a "generic" trust value from these comments, meaning that any seller's trust value that gets presented to a buyer is (a) presented to every other person who looks up that user, and (b) compiled from a system which is 99% positively biased. This is going to yield unnaturally positive trust scores in the system. Furthermore, eBay actually removes some negative comments from the system. As of September 2005, negative comments from users who have been in the system for less than 90 days get deleted from the system. [1]

We are proposing that trust values can be propagated throughout an e-commerce application between buyers and sellers, and that we can harness this information to compute a tailored trust value for a previously unseen user. A very basic example of this propagation might be as follows: Bob purchases from Mary and leaves a comment. Mary purchases from John and leaves a comment. If Bob's comment on Mary is positive, and Mary's comment on John is positive, then we might assume that if Bob were to purchase from Mary, there would be a positive comment. We make assumptions about the transitivity of trust in online applications. Section 2.1.2 outlines our arguments in detail.

[1]Correspondence to: John O'Donovan, Adaptive Information Cluster, School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: john.odonovan@ucd.ie

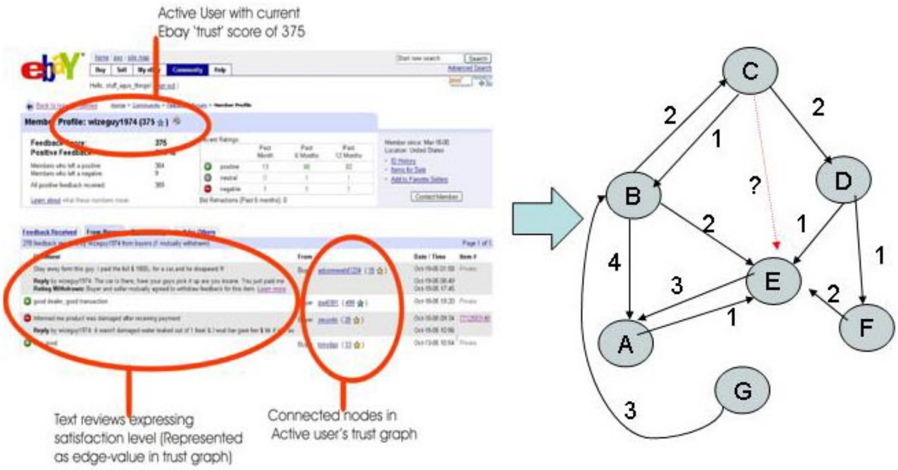

Figure 1. Personalized and Non-Personalized Trust Computation

Considering the fact that eBay is biased towards positive comments, we cannot use the existing eBay trust scores in our propagation mechanism. We examine several ways to compute trust-values which are different from the current eBay implementation, and introduce *AuctionRules*, an algorithm which we developed for this purpose. Details of the *AuctionRules* algorithm are given in section 3.1.

In our evaluation section we use a dataset of auction feedback comments which have been classified by real users in an online survey. These are used to test the classification accuracy of *AuctionRules* against 7 popular classification algorithms. Our results show that *AuctionRules* outperforms all 7 benchmark classification algorithms by an average of 21%, and achieves a false-negative rate of zero, compared with an average false-negative rate of 5% for the other algorithms. We also outline ongoing experiments for testing the accuracy of our trust propagation system on the same dataset.

Figure 1 illustrates the difference between the current non-personalized approach to trust-computation in eBay and our proposed personalized approach. Suppose user C wants to know how trustworthy user E is. Traditional (non-personalized) trust values are given by Equations 1 and 2 as a combination of all the incoming arcs. In these equations, n is the number of nodes used in the calculation. One of many computations for personalized trust is shown by Equation 3 as the combined trust of all the users in the path between nodes C and E. For example, Figure 1 shows the personalization in Equation 3 graphically in that node G will receive a trust value on node E through the connecting node B, whereas node C will receive a different trust value computed from the connecting nodes B and D. The benefit of this technique is that each node will receive a different, tailor-made trust prediction on user E, analogous to asking a friend about the seller in the local corner shop.

The main contributions of this paper are: firstly, the introduction of *AuctionRules*, an algorithm for extracting subtle indications of negativity from user-comments in online auctions, and secondly a trust-propagation mechanism which enables personalization of trust in online auctions.

$$Trust(E) = \frac{Trust(F,E) + Trust(A,E) + Trust(B,E) + Trust(D,E)}{n} \quad (1)$$

$$Trust(C,E) = Trust(E) \quad (2)$$

$$Trust(C,E) = \frac{Trust(C,B) * Trust(B,E) + Trust(C,D) * Trust(D,E)}{n} \quad (3)$$

2. Background Research

Background work for this paper is in two areas. Firstly, we examine the related work in the area of trust computation and transitivity in online auction sites. The second part of our background work involves a comparative survey of ten popular auction sites. This is predominantly an *applicability* survey, in which we examine how suitable our trust-propagation algorithm is to each auction site.

2.1. Related Work on Trust in Online Auctions

A large amount of research effort has focussed on issues of trust in e-commerce [15][14][2][5][17] and other online systems [10][11][9][4]. We will now examine work in formalising the concept of trust, it's transitive properties and it's use in e-commerce applications.

2.1.1. What is Trust?

A clear definition of what is meant by trust can be somewhat elusive. Marsh describes a formalized model of trust in [9]. This computational model considers both the social and the technological aspects of trust. Marsh defines some important categories for trust which are useful to our research. *Context-Specific Trust* arises when one user must trust another with respect to a specific situation. Marsh also defines *System Trust* as the trust that a user places in the system as a whole. Both of these concepts are especially relevant to our experiments with online auctions, in dealing with the individual users with whom we transact, and with the environment in which we make the transaction.

For our experiments with online marketplaces we must have a clear understanding of the differences between concepts of trust, and those of reputation. Resnick clearly differentiates between the concepts of trust and reputation in [14], citing the Oxford definition of reputation: *Reputation is what is generally said or believed about a person or things character or standing.* Resnick [14] sums up the difference in the following two plausible, normal statements. “I trust you because of your good reputation”. “I trust you despite your bad reputation”. Relating back to our previous example in Figure 1, *reputation* is given by Equations 1 and 2, whereas Equation 3 defines trust in the personalized sense given in [14].

2.1.2. Transitivity of Trust in Online Applications

Real world trust and reputation is propagated through friends and colleagues. [15] A potential stumbling block of this work lies in assumption that trust is transitive in the online world. For our experiments to be successful we require our trust values to have this property. There have been works which argue against this idea, for example Christianson shows in [2] that trust is not implicitly transitive. However, far more research supports transitive aspects of trust. Golbeck [4] introduces the *filmtrust* system, which operates successfully by propagation of trust. Similarly, the *moleskiing* system in [10] uses transitive trust. Experiments in [17] show trust propagation working successfully in *PeerTrust*, an experimental e-commerce application. Work in [11] and [5] also supports this concept in recommendation and e-commerce systems respectively.

Xiong and Liu [17] outline three basic trust parameters for online auction systems. 1: The amount of user satisfaction. 2: The *context*, defined as the number of previous transactions a user has been involved in. 3: A *balancing factor* of trust. (to offset false feedback). Using the *PeerTrust* system, a simulated network of 128 peers, they show that trust can be propagated with reasonable accuracy by using their three parameters.

Work by Jøsang et al. in [7] and [6] describe approaches to trust network analysis using subjective logic. They define a method for simplifying complex trust networks so they may be expressed in concise form, and use a formal notation for describing trust transitivity and parallel combination of trust paths. The core idea of this work is that trust can be represented as beliefs, and therefore be computed securely with subjective logic. The approach in [7] compares favourably with normalisation approaches such as Google's PageRank and the EigenTrust algorithm [8].

2.1.3. Trust on eBay.

Resnick highlights some relevant points which affect the current eBay reputation system in [14] and [15]. Buyers reputation matter less since they hold the goods until they are paid. Feedback can be affected by the person who makes the first comment, ie: feedback can be reciprocated. Retaliatory feedback and potential for lawsuits are strong disincentives for leaving negative comments. Anonymity is possible in eBay since real names are not revealed and the only thing validated at registration is an email address. Users can choose not to display feedback comments. Also "Unpaid Item" buyers cannot leave feedback [1], and users can agree to mutually withdraw feedback [1]. All of these points help to explain the lack of negative comments on eBay. Of course this does not mean that customers are satisfied. The eBay forums [1] highlight the fact that false advertising does occur on eBay. This should lead to more negative comments, but they are not being displayed. Xiong et al. [17] provide further reasoning for the imbalance of positive comments on eBay. Our proposed model of trust for eBay should provide a more realistic scale than the existing system.

2.2. Comparative Survey of Online Auction Sites

A survey was conducted of 10 online auction/retail sites. Full results of this survey are available on the web[2]. For each site we asked several specific questions relating to the potential applicability of a trust modelling and propagation module to the system. To assess whether or not our trust propagation system would be applicable to the site, we asked the following questions:

- What are the user roles in the system. *(Buyer/Seller/Both)*
- Is there an existing trust value? *(y/n)*
- If so, is it personalized *(y/n)*
- What is the percentage of positive comments?
- What are the review types *(Product/People/Both)*
- What are the requirements for making a rating *(Purchase/Registration/None)*
- What types of sale are provided *(auction/retail/both)*

Interestingly, none of the sites surveyed provided any personalization of trust scores. The trust propagation system required that users could be both buyers and sellers, comments were provided on transactions, and that the comments be mostly genuine. For this reason we enquired about the potential for malicious attacking [12] by assessing the cost of making a rating. We found that our application would be deployable on 7/10 of the surveyed sites. Users have the option of performing buyer or seller roles on 8/10 sites. On 6 of the sites, reviews were made about other users, and on 1 site reviews were on products only. 3 sites allowed reviews of both. An important statistic we gathered was that on average user feedback was over 94% positive across all the sites reviewed. This reaffirms the need for a negativity-capturing algorithm if we are to build a new trust

[1] http://forums.ebay.com
[2] http://www.johnod.net/research.jsp

model. *AuctionRules()* captures negativity in comments where users are dissatisfied with transactions but still provide positive ratings. This situation can arise for many reasons, such as fear of reciprocal negative comments as described by Resnick in [14]. Section 3.1 details this algorithm.

3. Building a Model of Trust

Figure 2 illustrates the process by which we build trust values based on rating data from the eBay site and from corresponding user-evaluations of those comments. Firstly a web crawler was designed to pull information from the live auction site. This system is easily generalisable to a broad range of auction sites, as long as comment information is available. We have a set of interfaces that can crawl data from several popular auction sites. For our experiments in this paper we use a subset of data from eBay. Crawled data is stored in an relational database in the basic form $(user_i, user_j, trust_{(i,j)})$, along with other information such as eBay trust score, number of transactions etc.

Some of the comments left by users will not contain much more information than the binary positive or negative rating that the commenter has given already. An example of such a comment is “Great Seller, Thanks!”. This comment does not provide us with any real information surplus to what we know already about this particular transaction. However, many of the comments on eBay do provide us with extra information which can be incorporated into the trust modelling process. Take the following for example: “Product delivered on time, perfect condition, nicely packaged, would buy from again!!”. This comment provides us a wealth of extra information about the seller, such as punctuality of delivery, package and product quality. These are some of the salient features of a trustworthy seller. It would require very advanced natural language processing techniques to fully analyse and understand every user comment on eBay. We have developed a technique for approximating the *goodness* of a user comment for the purposes of building our trust graph. The following section outlines this technique.

More restricted notions of trust are being examined. For example User A may be trusted in the context of book-recommendation, but not cars. Future work will show how a domain-constrained *PageRank* algorithm can produce trust values with more realistic transitive properties.

3.1. *The AuctionRules Algorithm*

AuctionRules operates under the assumption that people will generally use the same set of terms to express some form of dissatisfaction in their online auction comments. This has become apparent after initial manual examination of online auction comments, and subsequent automated testing. The algorithm captures negativity in comments where users have complained but still marked the comment as positive. This has been shown in [14] to occur a lot in situations where users are afraid of retaliatory negative comments.

AuctionRules is a machine learning algorithm. As with most ML techniques, training examples were required for the algorithm to learn. The algorithm works only with words and phrases which explicitly express negativity. Many of the words, expressions and characters in the raw comments were of no value to the learning process, so before training examples were compiled, preprocessing was done to reduce complexity. This algorithm does require some context information, in the form of short lexicons of special terms, and negative words. This information is *not* taken from, or specific to one site, and does work on *any* auction site with user feedback, so the context-dependency is very broad and the system widely applicable.

An implementation the “porter stemming” algorithm [13] was used to shorten comments. The standard porter stemmer uses many rules for removing suffixes. For example,

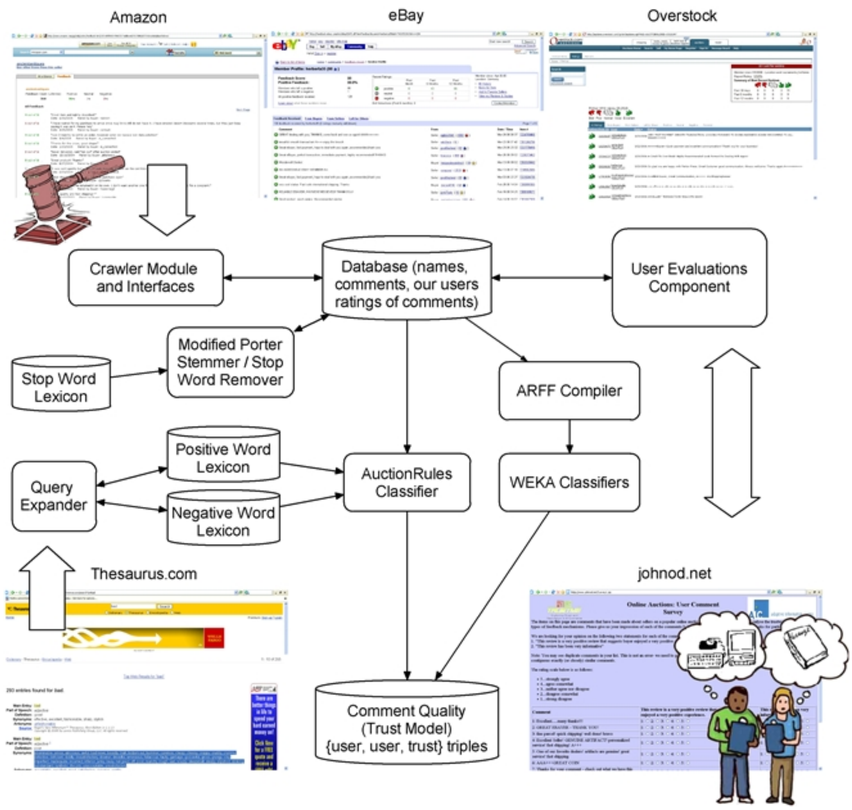

Figure 2. Graphical overview of the trust-modelling process. [Current implementation only uses eBay as a source for ratings.]

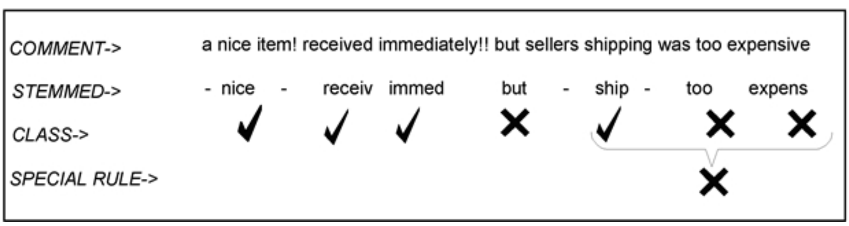

Figure 3. Stemming, stop-word removal and classification of a sample comment from eBay.

all of the terms in c are conflated to the root term *"connect"*. c = *connect*, *connected*, *connecting*, *connection*, *connections*. This reduces the number of terms and therefore the complexity of the data. The stemmer algorithm was modified to also stem characters not used by *AuctionRules*, such as "?, !, *, (,)" for example.

Data complexity is further reduced by removal of stop-words. Initially, Google's stop-word list[3] was removed. This list was embellished with over 50 frequent words manually selected from user comments (on *different* auction sites) deemed to be of no help in ascertaining negativity. Examples of words in this list are: "Product, Item, Seller, Those, These, Book, Buyer". In a sample test of 4065 tested terms, there were only 556 unique after stemming/stop word removal. This is one unique term in every 7.3, which is a large overlap. When *AuctionRules* works in probabilistic mode, it will output the number of

[3]http://www.ranks.nl/tools/stopwords.html

bad terms in the comment, which is 3 in this example, over the maximum number of bad terms found in a comment, which was 5 in our tests, meaning that the probability of the example comment being negative is 60%.

AuctionRules starts as a majority class predictor over the training data. Negative comments are isolated using lexicon of single negative terms such as “bad”, “awful” etc, and a set of checks for negative prefixes such as “not” before positive terms such as “good”. Each term in the comment is tested individually, with the preceding term where possible. A *hit* arises when a negative term or phrase is found in the comment. A threshold value is used to determine the number of negative hits required to sway the classifier. Figure 3 shows *AuctionRules* working on a sample comment from eBay. In this particular example, *AuctionRules* will always classify the comment as *negative* because it triggers a special rule. There are a number of such rules used in the algorithm, e.g. when the stemmed term “‘ship” is found with any term from a term-list indicating “costly”, the comment is classified as negative. These rules are manually defined from examining user-comment data over a range of sites. They are completely independent of the collected data. Currently the rules are simple and there are cases that will be misclassified, although these are rare. One such example is: "the ship is anything but bad". The algorithm is currently being improved to handle more complicated cases.

To provide training examples for the algorithms we had to manually classify the data. User’s opinions were collected on each comment in an online evaluation. Details of this evaluation are given in the following section. Comments were rated on a scale of 1 to 5 (1 = strong negative). A threshold value of 2 was chosen as the pivotal point for a positive comment by the *AuctionRules* algorithm. This value was chosen because it produced the highest accuracies when we varied the threshold for a positive comment over different runs of several popular classifiers in pre-testing. Any comment rated 2 or lower was tagged as negative, and the rest positive.

Figure 2 also shows a query expansion module, which is un-tested at present. This module queries an online thesaurus[4] for a list of synonyms, and expands each individual term *AuctionRules* is testing to include the list of synonyms. Each word in this new list is then compared to the lexicon of negative (and/or positive) words. When this expansion is used there are many false hits, so the threshold for the number of hits to sway the classifier is greatly increased. Experiments to test this technique in more detail will be explained in a future paper.

3.2. Live User Evaluations of eBay Comments

From the user evaluations[5] 1000 ratings on eBay comments were collected. In this survey, users were asked to rate the positiveness of each comment on a Likert scale of 1 to 5. 10 comments were presented to a user in each session. Each comment was made by different buyers about one seller. Users were required to answer the following:

- How positive is the comment (Average rating: 3.8442)
- How informative is the comment (Average rating: 3.1377)
- Would you buy from this seller (Average rating: 4.0819)

In the current implementation of *AuctionRules*, only results from the first question are used to develop training examples. For future experiments we may incorporate results from the other questions. Permission was sought from eBay inc. to use the information from the eBay website in our experiments. We had to gather the data ourselves by using a specially tailored web crawler.

[4]http://thesaurus.reference.com
[5]www.johnod.net/Survey1.jsp

3.3. Classification of Comments in WEKA

In our experimental evaluation we show results of a range of classification algorithms running on our pre-classified set of feedback comments. We used the Weka corpus [16] of machine learning algorithms to run most of the experiments. To use the Weka classification algorithms from their system, data must be input in "ARFF" format. An *ARFF-Compiler* program was written to automatically feed the comments into the Weka system. Figure 2 shows where this fits into the architecture. Each comment was tagged with its classification from the user evaluation. The header files consisted of a list of attributes used by Weka's algorithms. The attribute list for our data is a list of comma separated *unique* terms across the entire set of comments. Details of the Weka experiments compared with *AuctionRules* are shown in the evaluation.

AuctionRules can be set to output binary or probabilistic classifications of comments, based on the number of negative terms. As shown in Figure 2, output from the algorithm is a set of triples of the form: $e_{(i,j)} = (user_i, user_j, trust_{(i,j)})$. This forms our basic units of trust to be used in the propagation mechanisms described in the following section.

4. Representing the Auction as a Trust Graph

So far we have explained the *AuctionRules* algorithm generates a broader scale of trust than the current feedback we see on ebay by drawing on information in user comments. Figure 6.2 in the evaluation section shows this improvement. The new trust values are of the form $e_{(i,j)} = (user_i, user_j, trust_{(i,j)})$. Now we explain how the model generated by *AuctionRules* enables us to generate personalized trust values.

We construct a trust-graph of our subset of the eBay marketplace so that we can find paths between buyers and sellers. There were a number of factors to consider when constructing this graph. The three scenarios in Figure 4 show these considerations. In the graph an arrow represents some unidirectional trust value. On eBay, users can play three different roles: buyer, seller or both. This creates different communication patterns between users with different roles. For example, Resnick explains in [14] that seller trust matters less since they hold the product until payment is made. When a transaction happens on eBay, a buyer leaves a comment on a seller and vice-versa. We could consider only buyers comments on sellers, since the bulk of the hazard in the transaction lies with trusting the *sellers* [14]. Figure 4 (a) depicts this approach. However, as the diagram shows, there is much less connectivity in the resulting graph. In fact, the only way for a buyer node r can *know about* a seller p is through a node q that performs both roles, and has interacted with r and p.

To overcome this constraint, we assume all nodes to be of type q, that is, they can all be both buyers and sellers. There is no distinction drawn between comments from buyers and those from sellers. This simplification allows much greater connectivity in the trust graph. Every node in the resulting graph will have an *even* number of edges entering and leaving it. Furthermore, there will be an *equal* number of edges entering and leaving the node. Figure 4 (b) shows this graphically. In many cases there are multiple transactions between the same buyer and seller, as shown in Figure 4 (c). When this occurs, we take the average value over each linking edge in the graph.

5. Generating Personalized Trust

Figure 5 shows the implementation of the personalization stage of our system. An ID value for the user seeking the trust value is passed to the system, along with an ID for the potential seller or buyer that the user is querying. The system queries the database of trust values and returns the two shortest paths between the two nodes. A path between two

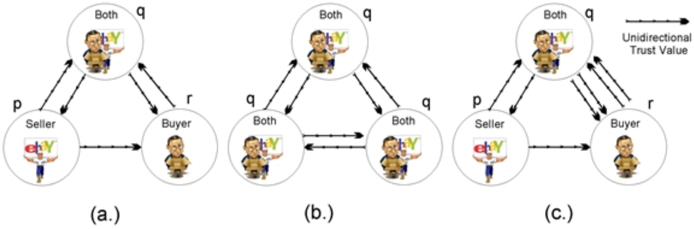

Figure 4. Graphical depiction of possible interactions between buyers and sellers in the eBay marketplace

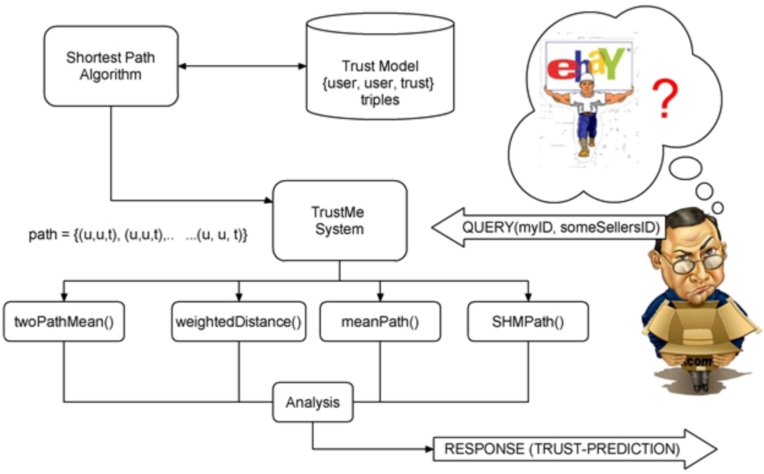

Figure 5. Graphical overview of the trust prediction phase of the system. [Some of the analysis modules are not included in the illustration.]

users is represented in the form $P = (u_{source}, u, t), (u, u, t), ...(u, u_{sink}, t)$ Where u_{source} is the user seeking a trust value t on user u_{sink}. The system then combines the trust values along these paths in four different ways. These combinations are given below. The system can be set to use any of the four techniques for combining trust scores along the paths.

- *weightedDistance* - The average trust score over all the edges in the shortest path, discounted by the distance from the source.
- *meanPath* - The average trust score over all the edges in the shortest path between the source node and target node.
- *twoPathMean* - The average of the *meanPath* of the shortest path in the graph and the *meanPath* of the second shortest path.
- *SHMPath* - The simple harmonic mean of the trust scores over all edges in the shortest path.

6. Experimental Evaluation

Ideally our data should have a high level of overlap between buyers and sellers, as well as rich textual comments on each transaction made. eBay provided a structured way for users to leave good textual comments on their sales and purchases. This domain has millions of users so we chose a subset which had very high overlap amongst its users. This subset was the purchase and sales of Egyptian Antiques, because there were a large number of users who played the roles of both buyer and seller. In cases where the trust

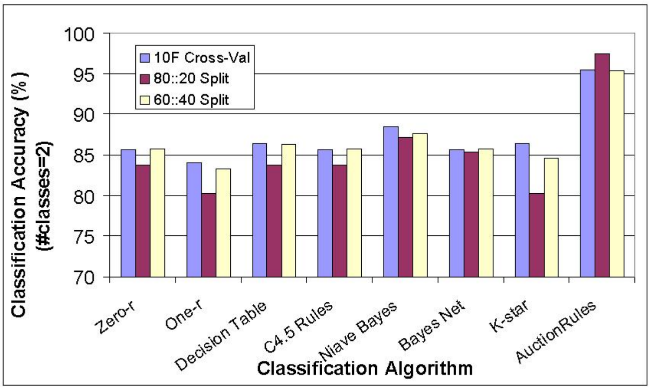

Figure 6. Classification Accuracy [classification distribution from user evaluations: 36% positive, 63% negative, using a threshold of 4 or higher for a positive comment.]

graph sparse, or not connected at all, the system can present trust for a node simply by averaging the incoming trust values for that node.

Initially we crawled over 10,000 comments from the eBay site. For the following experiments, we used only the set of comments which were rated by real people in our user evaluations. This is a set of 1000 classified user comments. From this set we found that on average there were 5.08 terms per comment, the max number of terms in a comment was 16. The set consists of entries from 313 buyers on 14 sellers.

6.1. Preliminary Experiment 1: Comparing Classifier Accuracy for Computing Trust

We needed to assess how well the system extracted trust values from raw textual comments. To examine this empirically, the classification accuracy of the *AuctionRules* algorithm was tested against 7 popular algorithms. We chose three rule-based learners, *Zero-r, One-r* and *Decision Table*, a tree learner *C4.5 rules*, two Bayes learners, *Naive Bayes* and *BayesNet*, and a lazy learning algorithm *K-Star*.

Figure 6 shows results of this experiment. For each algorithm we performed three runs. a 60:40 train-test split, an 80:20 split, and a 10-fold cross validation of the training set, which randomly selects a training set from the data over 10 runs of the classifier and averages the result. In the experiment each algorithm made a prediction for every value in the test set, and this prediction was compared against the training set. *AuctionRules* beat all of the other classifiers in *every* test we performed, achieving over 90% accuracy in all of the evaluations, 97.5% in the 80:20 test, beating the worst performer *K-Star* by 17.5%, (relative 21.2%) and it's closest competitor *Naive Bayes* by 10.5%, giving a relative accuracy increase of 12.7%.

In addition to numerical accuracy, we examined where the high accuracy results were coming from more closely by assessing the confusion matrix output by the algorithms. This was necessary since prediction of false negatives would have an adverse effect on the resulting trust graph. This phenomenon has been discussed by Massa in [10] with respect to the *Moleskiing* application, and Golbeck in [4] with respect to the TrustMail application. Figure 6.1 shows *AuctionRules* outperforming all of the other algorithms by predicting no false negatives. All of the algorithms displayed similar trend to the ones in Figure 6.1, which shows results of the 80:20 classification experiment which had a test set of 234 comments.

	AuctionRules		*NaiveBayes*		*Decision Table*		*One-r*	
	+'ve	**-'ve**	+'ve	**-'ve**	+'ve	**-'ve**	+'ve	**-'ve**
+'ve	91.4	*0*	84.1	1.2	84.6	1.2	77.3	8.1
-'ve	4.7	4.7	11.1	2.9	12.3	1.7	8.5	5.9

Table 1. Confusion matrices showing percentage true and false negatives and positives for four of the algorithms tested. [All of the other algorithms had similar results to the ones displayed.]

6.2. Preliminary Experiment 2: Trust Distributions

As we mentioned in the introduction, there are too many positive comments on online auction sites such as eBay. Our *AuctionRules* algorithm performs accurate classifications when compared against classifications of real people.

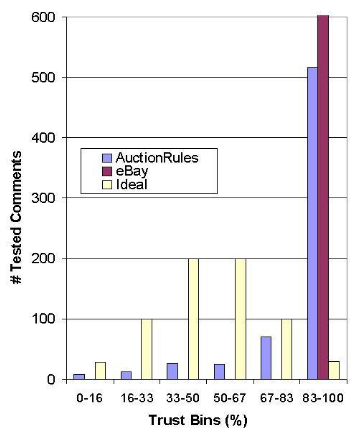

Figure 7.: Comparison of distributions between current eBay trust values and *AuctionRules* generated values.

Now we must ask what effect this will have on the resultant trust values, especially in comparison with the existing values on eBay. Figure 6.2 shows the distributions of trust values produced by *AuctionRules* compared to the existing eBay trust values, and with what we believe, based on our distribution research in [11] to be the ideal standard for trust distribution in an online auction, where the model can isolate small numbers of highly trustworthy and similarly highly untrustworthy users. *AuctionRules* was set to output probabilistic classifications which were used as trust values for the distribution graphs. This was done by counting the number of negative terms in a comment and then dividing by the max number of negative terms found in a comment to produce normalized values. Figure 6.2 shows that the vast majority of the computed trust scores are still highly positive. This may not look like a large improvement at first glance, but considering that most of the comments in our data genuinely were positive, and that *AuctionRules* takes about 16.6% of the eBay trust values from the 83-100 bracket and distributes them across the entire scale, we can see that this is a positive result for the algorithm.

6.3. Proposed Experiment: Accuracy of the trust propagation model

This experiment is designed to test accuracy of the propagated trust values, i.e. ability of the system to use the trust graph to predict a trust value for a new node. On eBay, this equates to the systems ability to provide a personalized trust value of a seller to a potential buyer who has no previous interaction with that seller. In this paper we are only mentioning this ongoing experiment, in which we temporarily remove each edge in the trust graph and use our prediction techniques from Figure 5 make trust predictions for the missing edge. The full set of results from this experiment will be presented in a later paper.

7. Conclusions and Future Work

We have proposed that the current trust scoring system on eBay is massively biased towards positive comments, which tend to generate more revenue for the system. [1]. We have discussed related publications that back up this statement. [14][15].

As an alternative/addition to the current systems we have introduced our ideas for a trust modelling system for online auctions, using eBay as an example. In our proposed system, numerical trust values are mined directly from user comments using *Auction-*

Rules, a new classification algorithm which captures negativity in user comments. Our system also facilitates the propagation of the trust values throughout the social network that is formed by the eBay marketplace. Propagation of trust values allows us to compute a *personalized* trust score on a prospective seller to be presented to a buyer. In our background work we have carried out a comparative survey of ten popular online auction sites, which determined that our application would be deployable on seven of the ten sites.

In our evaluation section we outlined two initial experiments to test the validity of our system. Firstly to test the accuracy of the trust mining algorithm, and a second experiment to examine the relative distributions of the existing trust rating system against our computed trust values. Results show a more realistic distribution using the *AuctionRules* values, and consistent improvements of up to 21% over seven popular classification algorithms. *AuctionRules* also produces a false negative rating of 0% compared with 8.1% from other tested algorithms.

Ongoing research includes testing the accuracy and coverage of the trust-propagation mechanisms, testing the query expansion module, incorporating new data from other auction sites, and refining the *AuctionRules* algorithm to produce higher accuracy.

8. Acknowledgements

This material is based on works supported by Science Foundation Ireland under Grant No. 03/IN.3/I361, and in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152.

References

[1] Brian Burke. Ebay town-hall meeting. Transcript from Senior Manager. eBay Marketplace Rules and Policy, eBay Inc, San Jose, CA. Feb 28th 2006. available at http://pics.ebay.com/aw/pics/commdev/TownHallTranscript_022806.pdf.

[2] Bruce Christianson and William S. Harbison. Why isn't trust transitive? In *Proceedings of the International Workshop on Security Protocols*, pages 171–176, London, UK, 1997. Springer-Verlag.

[3] eBay Incorporated. Fourth quarter and full year financial results for 2005. Press release, San Jose, CA. Jan 18 2006. available at http://investor.ebay.com/news/Q405/EBAY0118-123321.pdf.

[4] Jennifer Golbeck and James Hendler. Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *Proceedings of EKAW'04*, pages LNAI 2416, p. 278 ff., 2004.

[5] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust, 2004.

[6] Audun Jøsang, Elizabeth Gray, and Michael Kinateder. Analysing Topologies of Transitive Trust. In Theo Dimitrakos and Fabio Martinelli, editors, *Proceedings of the First International Workshop on Formal Aspects in Security and Trust (FAST2003)*, pages 9–22, Pisa, Italy, September 2003.

[7] Audun Jøsang, Elizabeth Gray, and Michael Kinateder. *Simplification and Analysis of Transitive Trust Networks*. *Web Intelligence and Agent Systems: An International Journal*, pages 1–1, September 2005. ISBN ISSN: 1570-1263.

[8] Sepandar Kamvar, Mario Schlosser, and Hector Garcia-Molina. The EigenTrust Algorithm for Reputation Management in P2P Networks. In *Proceedings of WWW2003*. ACM, 2003.

[9] S. Marsh. Formalising trust as a computational concept. *Ph.D. Thesis. Department of Mathematics and Computer Science, University of Stirling,1994*.

[10] Paolo Massa and Bobby Bhattacharjee. Using trust in recommender systems: an experimental analysis. *Proceedings of 2nd International Conference on Trust Managment, Oxford, England*, pages 221–235, 2004.

[11] John O'Donovan and Barry Smyth. Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 167–174. ACM Press, 2005.

[12] John O'Donovan and Barry Smyth. Is trust robust? an analysis of trust-based recommendation. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, pages 101–108. ACM Press, 2006.

[13] M. F. Porter. An algorithm for suffix stripping. *Readings in information retrieval*, pages 313–316, 1997.

[14] Paul Resnick and Richard Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce. Volume 11 of Advances in Applied Microeconomics.*, December 2002.

[15] Paul Resnick, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara. Reputation systems. *Communications of the ACM*, December 2000. Vol. 43, No. 12.

[16] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, 2005.

[17] Li Xiong and Ling Liu. Building trust in decentralized peer-to-peer electronic communities. In *Fifth International Conference on Electronic Commerce Research (ICECR-5)*, 2002.