The Sound of Silence: Mining Implicit Feedbacks to Compute Reputation

Mikołaj Morzy[1] and Adam Wierzbicki[2,⋆]

[1] Institute of Computing Science
Poznań University of Technology
Piotrowo 2, 60-965 Poznań, Poland
Mikolaj.Morzy@put.poznan.pl

[2] Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa, Poland
adamw@pjwstk.edu.pl

Abstract. A reliable mechanism for scoring the reputation of sellers is crucial for the development of a successful environment for customer-to-customer e-commerce. Unfortunately, most C2C environments utilize simple feedback-based reputation systems, that not only do not offer sufficient protection from fraud, but tend to overestimate the reputation of sellers by introducing a strong bias toward maximizing the volume of sales at the expense of the quality of service.

In this paper we present a method that avoids the unfavorable phenomenon of overestimating the reputation of sellers by using implicit feedbacks. We introduce the notion of an implicit feedback and we propose two strategies for discovering implicit feedbacks. We perform a twofold evaluation of our proposal. To demonstrate the existence of the implicit feedback and to propose an advanced method of implicit feedback discovery we conduct experiments on a large volume of real-world data acquired from an online auction site. Next, a game-theoretic approach is presented that uses simulation to show that the use of the implicit feedback can improve a simple reputation system such as used by eBay. Both the results of the simulation and the results of experiments prove the validity and importance of using implicit feedbacks in reputation scoring.

1 Introduction

Internet economy is doing very well. According to eMarketer, the e-commerce market is steadily growing with annual gains reaching 25% in 2004 and 21% in 2005. The growth is broad-based and distributes almost equally among all categories of retail, travel and entertainment. Projection for the future is optimistic: annual gains will continue to grow at a double-digit level reaching retail revenues

⋆ The research was supported by the Polish Ministry of Science and Higher Education under grant 3T11C 005 27 "Models and Algorithms for Efficient and Fair Resource Allocation in Complex Systems."

P. Spirakis et al. (Eds.): WINE 2006, LNCS 4286, pp. 365–376, 2006.

of $139 billion in 2008 in the United States alone. Online auctions are among the most popular and important e-commerce services. It is estimated that over 15% of all e-commerce sales can be attributed to online auctions. EBay, the global leader in online auctions, has over 56 million active users (and over 95 million registered users). Annual transactions on eBay surpass $23 billion, with approximately 12 million different items posted simultaneously on eBay at any point in time and slightly less than 1 million transactions committed daily. This immense marketplace for customer-to-customer e-commerce provides means for anonymous and geographically dispersed users to seal retail transactions. But an important question arises: how does one estimate the reputation of an anonymous business partner and how does one develop trust when there is hardly any history of business contacts between any two partners?

A reliable reputation system is crucial for enabling a fair and credible environment for e-commerce activities. The quality of the reputation system directly affects the credibility of an online auction service and impacts the amount of fraud present on the online auction market. Unfortunately, fraud is still the main reason hindering further development of online auctions. According to National Fraud Information Center, online auctions account for 42%[1] of all registered complaints with an average loss of $1,155. The number of complaints grows quickly (12,315 complaints in 2005 compared to 10,794 in 2004) as well as the total loss ($13,863,003 in 2005 compared to $5,787,170 reportly lost in 2004). Online auction fraud definitely outranks other popular types of scams. Therefore, the reputation system used by an online auction site must be robust enough to safeguard the online community of auction participants against fraudsters.

Devising a robust and fraud-free reputation system is difficult for various reasons. Most importantly, the reputation system must take into consideration high asymmetry between buyers and sellers in online auctions. These two classes of auction participants are exposed to different types of risk. Sellers are almost never threatened financially, because they can postpone the shipment of the merchandise until the payment is delivered. Therefore, sellers are generally not concerned with the reputation of their business partners. On the other hand, buyers decide upon participation in an auction solely based on the reputation of a seller. Furthermore, after delivering the payment buyers are still in hazard of receiving no merchandise, or receiving merchandise of lower quality and inconsistent with the initial offer. From a buyer's point of view, a credible estimation of seller's reputation is indispensable for secure and successful trade.

Most online auction sites use a simple feedback-based reputation system [10]. Typically, parties involved in a transaction mutually post feedbacks after the transaction is committed. Each transaction can be judged as 'positive', 'neutral', or 'negative'. The reputation of a user is simply the number of distinct partners providing positive feedbacks minus the number of distinct partners providing negative feedbacks. As pointed out in [5], such simple reputation system suffers from numerous deficiencies, including the subjective nature of feedbacks and the

[1] This number is grossly underestimated due to eBay's reluctance to cooperate with NFIC. NFIC estimates that the real number is closer to 70%.

lack of transactional and social contexts. We identify yet another drawback of feedback-based reputation systems: these systems do not account for psychological motivation of users. Many users refrain from posting a neutral or negative feedback in fear of retaliation, thus biasing the system into assigning overestimated reputation scores. This phenomenon is manifested by high asymmetry in feedbacks collected after auctions and, equally importantly, by high number of auctions with no feedback provided. We believe that many of these missing feedbacks convey implicit and unvoiced assessments of poor seller's performance which must be included in the computation of seller's reputation to provide an unbiased estimation of seller's reliability.

In this paper we introduce the concept of an implicit feedback. Implicit feedback is a useful, actionable pattern hidden in large amounts of online auction data. We mine the history of user feedbacks to discover missing feedbacks that were left out purposely and we include these implicit feedbacks in reputation scoring. We present an efficient and flexible strategy for identifying implicit feedbacks and we compare it to a simple majority voting strategy. We present a twofold experimental evaluation of our proposal using both game-theoretic simulation and experiments involving a large set of real-world data. The results of conducted experiments clearly indicate an important impact of using implicit feedbacks in reputation scoring. The paper is organized as follows. In Sect. 2 we present the related work on the subject. The existence and computational feasibility of implicit feedback are presented in Sect. 3. Two strategies for discovering implicit feedbacks are also presented. Section 4 contains the results of the experimental evaluation of our proposal. In this section we show how a simple reputation algorithm, such as used by eBay, can be greatly enhanced by using the implicit feedback. The paper concludes in Sect. 5 with a summary and future work agenda.

2 Related Work

Reputation systems [11] are trust management systems that are used to enable trust between anonymous, heterogeneous, and geographically dispersed business partners. One of the most important tasks of reputation systems is to provide an economical incentive to behave honestly [2] and to reward participants for fair behavior [7]. Among many definitions of trust, the one that is the most useful in the context of online auctions is that trust is a subjective expectation that other agents (i.e., buyers and sellers) will behave fairly [3,9]. In this context, fair behavior is defined as carrying out the agreed transaction to the best of agent's ability. This definition of trust relates reputation systems to the theory of justice that can be used to evaluate reputation systems. A reputation system works well if all agents that behave fairly receive just payoffs.

One of the most well-known method to evaluate justice is the use of the Lorenz curve and the Gini coefficient [4,6]. The Lorenz curve is a graphical representation of the ordered cumulative distribution function of a distribution of goods (income, resources, etc.). Consider a set of n agents that receive shares of goods denoted by $x_1, \ldots, x_n$. First, we need to sort the shares of agents incrementally, receiving a permutation of shares $x_{i_1}, \ldots, x_{i_n}$. From the resulting permutation we take cumulative sums $\theta_i = \sum_{j=1}^{i} x_{i_j}$. These sums are sometimes normalized by the sum of all shares, θ_n. The values of θ_i plotted against i form the Lorenz curve. Note that if all shares x_i are equal, then the Lorenz curve will be a straight line. For unequal distributions, the Lorenz curve is convex. An idealized perfect distribution of goods is given by $\theta_i = (i * \theta_n/n)$. The area between the Lorenz curve and this perfect distribution line (normalized by $2\theta_n$) is known as the Gini coefficient. We shall use the Gini coefficient as the measure of justice for the evaluation of reputation system effectiveness.

3 Existence of Implicit Feedback

A close investigation of the distribution of feedbacks reveals a striking deviation. The examined dataset contains data on a sample group of 10 000 buyers collected over the period of six months. There are 656 376 committed auctions and 890 876 mutual feedbacks. Table 1 summarizes data statistics.

Table 1. Distribution of feedbacks

	negative	%	neutral	%	positive	%	$\sum$	%
buyer	4318	1%	2877	0.6%	445 723	98.4%	452 918	69%
seller	2558	0.6%	553	0.1%	434 847	99.3%	437 948	66%

Buyers provided 452 918 feedbacks, which accounts for 69% of all examined auctions. Note that over 30% of all auctions are not sealed with a feedback. Almost all registered feedbacks are positive (98.4%), with only 1% of negative and 0.6% neutral feedbacks. Similar characteristics can be observed for feedbacks provided by sellers, although sellers are slightly less eager to provide a feedback in general. Similar results are reported in [12], so we believe that such distribution is quite typical for online auction sites. Table 1 presents a grossly optimistic view of the quality of service offered by participants. There are two interesting points to make. First, neutral feedback is missing, the scope for positive feedback ranges from an open praise to the acknowledgement of a correct auction (but nothing more), and negative feedback appears only when the quality of service becomes totally unacceptable. Second, more than 30% of auctions did not finalize with a feedback. In many of these auctions sellers conducted poorly, but the quality of service was either bearable, or the buyer was intimidated and afraid of a retaliatory negative feedback. In both cases the reputation of a seller should be affected negatively. We refer to these purposely omitted feedbacks as implicit feedbacks that indicate a seller's performance that is unsatisfactory and not deserving a praise, yet passable. Listening to these silent, unvoiced feedbacks makes the reputation estimation more credible. Alas, current reputation systems are not aware of the existence of implicit feedbacks and do not incorporate implicit feedbacks into reputation scoring.

Not every missing feedback should be regarded as an implicit assessment of user's performance. A feedback might be missing for various reasons, e.g., one of the trading parties might be an unexperienced user who does not know how to post a feedback. One simple strategy is to check the history of user's feedback and compute the ratio of user's auctions for which a given user has posted a feedback. If the majority of user's auctions have been sealed with a feedback, a missing feedback for a given auction might indicate a purposeful omission of the feedback, i.e., an implicit feedback. We call this strategy the *majority voting strategy.* We also devise a more complex and flexible *cosine strategy* presented next. Let $F(u_i) = \langle f_1, f_2, \ldots, f_n \rangle, f_i \in \{0,1\}$ be a chronologically ordered list of feedback flags posted by the user u_i, where $f_k = 0$ denotes the fact that the user u_i did not provide a feedback for her k-th auction, and $f_k = 1$ denotes the fact that the user u_i explicitly provided a feedback for her k-th auction. We arbitrarily assume that the effect of each auction experience (either positive or negative) influences the next two auctions of a given user[2]. $F(u_i)$ can be transformed into an ordered list of trigrams $T(u_i) = \langle t_1, t_2, \ldots, t_{n-2} \rangle$, where $t_i = f_i f_{i+1} f_{i+2}$ is a binary concatenation of feedback flags for the i-th auction with feedback flags for the consecutive two auctions. There are $2^3 = 8$ possible trigrams represented by binary numbers ranging from 000 (three consecutive auctions do not have a feedback) to 111 (three consecutive auctions have a feedback). Thus, $T(u_i)$ can be represented as a vector $\bar{T}(u_i) = \left[t_i^0, \ldots, t_i^7\right]$, where t_i^n is the number of occurrences of the n-th trigram in $T(u_i)$. We perceive $\bar{T}(u_i)$ as a condensed representation of feedback habits of the user u_i. Having transformed the original history of user feedbacks into an 8-dimensional vector we can compare this vector to a template vector representing a user who almost never provides a feedback for her auctions (in our experiments we have used the template vector $\bar{T}(0) = [1, 0.1, 0.1, 0.01, 0.1, 0.01, 0.01, 0]$, where three consecutive auctions without a feedback have the weight 1, two missing feedbacks have the weight 0.1, and one missing feedback has the weight 0.01). Let k-th auction of the user u_i does not have a feedback. First, we build $F(u_i) = \langle f_1, f_2, \ldots, f_k \rangle$, which is transformed into $T(u_i) = \langle t_1, t_2, \ldots, t_{k-2} \rangle$, and the resulting list $T(u_i)$ is transformed into the vector $\bar{T}(u_i)$. Next, we compute the Ochini coefficient (the cosine similarity function) between $\bar{T}(u_i)$ and $\bar{T}(0)$ as follows

$$Ochini(\bar{T}(u_i), \bar{T}(0)) = \frac{\sum_{k=0}^{7} t_i^k * t_0^k}{\sqrt{\sum_{k=0}^{7} (t_i^k)^2 * \sum_{k=0}^{7} (t_0^k)^2}}$$

If $Ochini(\bar{T}(u_i), \bar{T}(0)) < \beta$, where β is a user-defined threshold, we conclude that the two vectors are similar and the omission of a feedback should not be regarded as an implicit feedback.

Example 1. Let us assume a user u with the following list of feedback flags: $F(u) = \langle 0, 1, 0, 1, 1, 0 \rangle$. The user u participated in six auctions and did not pro-

[2] The authors acknowledge that the choice of three consecutive auctions as the range of psychological influence of an auction outcome is arbitrary and the correctness of this assumption remains open for discussion.

vide feedback for three of them. We want to know if the last missing feedback is a purposeful omission. First, the list of feedback flags $F(u)$ is transformed into a list of trigrams $T(u) = \langle 010, 101, 011, 110 \rangle$. Next, the list of trigrams is transformed into a compact vector representation $\bar{T}(u) = [0, 0, 1, 1, 0, 1, 1, 0]$. The final result is $Ochini(\bar{T}(u), \bar{T}(0)) = 0.09$. After a certain period of time the user u participates in more auctions and gains experience. Let us assume that, after a while, the list of feedback flags for the user u is $F(u) = \langle 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0 \rangle$. We want to decide on the last missing feedback as being an implicit feedback. The list of trigrams is $T(u) = \langle 010, 101, 011, 110, 101, 011, 111, 110, 101, 011, 110 \rangle$ and the vector representation is $\bar{T}(u) = [0, 0, 1, 3, 0, 3, 3, 1]$. Now the computation of the Ochini coefficient yields $Ochini(\bar{T}(u), \bar{T}(0)) = 0.035$. As can be seen, this procedure is flexible and allows for temporal changes in feedback habits.

To prove the existence of the implicit feedback we begin by investigating the distribution of the number of missing feedbacks per user (in this experiment we include only buyers). The results of the experiment are depicted in Figure 1. Interestingly, there are only a few buyers with more than 20 missing feedbacks. This might indicate that most of the missing feedbacks are in fact purposeful omissions, thus turning the missing feedbacks into implicit feedbacks. When we have constrained our search to buyers who had participated in at least 10 auctions, the average percentage of missing feedbacks dropped to 11.6%, which indicates that experienced users are even less likely to omit a feedback.

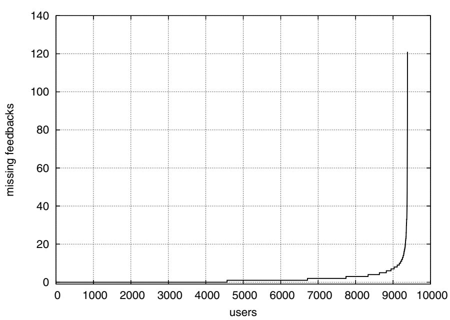

Fig. 1. Missing feedback distribution

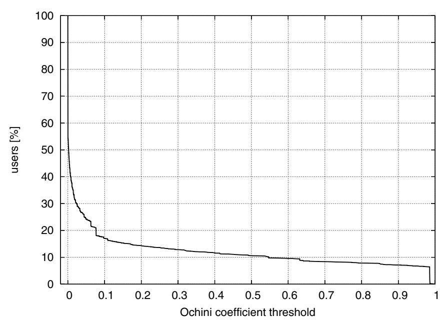

Fig. 2. Ochini coefficient selectivity

Figure 2 presents the user selectivity depending on the value of the Ochini coefficient. Recall that the Ochini coefficient represents the similarity between a given user's feedback vector and the template vector of a hypothetical 'I-don't-do-feedbacks' user, with the values closer to 1 representing high similarity and the values closer to 0 representing high dissimilarity. The figure presents the percentage of users who would be considered as generally not providing feedbacks, given the value of the Ochini coefficient threshold. For reasonable values of the Ochini coefficient threshold (i.e., 0.5 and above) less than 10% of buyers are regarded as reluctant to provide feedbacks, which means that their missing feedback would not be considered as implicit feedback. Again, this result proves that for the majority of buyers a missing feedback is an important, yet unvoiced, assessment of business partner's performance.

4 Effectiveness of Using Implicit Feedback

To evaluate a reputation system it is necessary to find out how this system affects the behavior of users and the outcome of user transactions. Ideally, we would like to know whether a reputation system enables trust: all honest users should trust other honest users and should be treated fairly by other honest users. On the other hand, all dishonest users should not be trusted and therefore should not participate in transactions. In this section we compare a simple reputation algorithm, such as used by eBay, to a more complex algorithm that uses implicit feedback. Section 4.1 presents the design of the simulator of online auctions. The results of conducted simulations are reported in Sect. 4.2. The impact of implicit feedback on real-world data is presented in Sect. 4.3.

4.1 The Simulator

Prior to starting the simulation we had to make a decision about a sufficiently realistic, yet not too complex model of the auctioning system, of user behavior, and of the reputation system. We choose to simulate the reputation system almost totally faithfully, the only simplification is that we use only positive and negative feedbacks. The behavior of a user is also realistic: the user takes into account the reputation when choosing a business partner. The user also decides whether she wishes to report or not, depending on the type of report. Users can cheat in reports, as well as in transactions. Users may also use transaction strategies that depend on the history of their individual interactions with other participants.

The auctioning system, on the other hand, has been simplified. We reflect that the simulation of the entire auction process is unnecessary. Rather, we simulate the selection of users using a random choice of a set of potential sellers. The choosing user (i.e., the buyer) selects one of the sellers from among users with the highest reputation in the set. The auction itself has also been simplified. We use a popular game-theoretic model of an auction, namely, the iterated Prisoner's Dilemma [1].

In the simulator, a number of agents that represent users interact with each other. Each interaction represents an auction between a seller and a buyer. The reputation system is maintained by a reputation server that is also used to summarize the outcomes of agent interactions. Each agent is characterized by the following parameters: r^+, the probability that an agent will send a report if it is positive, r^-, the probability that an agent will send a report if it is not positive, the chosen game strategy, the reputation threshold ρ_{min} that is used by some strategies, and the probability of cheating c, that is used by some strategies. We can specify the number of agents and every agent can have distinct parameters. However, we usually partition all agents into two sets that have the same parameters, called the honest and dishonest agents.

The two game strategies used in the simulations are: to cheat with the probability c or to play Tit-for-Tat with a reputation threshold ρ_{min}. Tit-for-Tat is a famous strategy for the iterated Prisoner's Dilemma game. This strategy works simply by repeating the move made by the other agent in the previous encounter. If two agents meet for the first time, the classic Tit-for-Tat strategy forces the agents to cooperate, thus allowing the agents to start an unending pattern of honest transactions. We modify Tit-for-Tat to use a reputation threshold: if two agents meet for the first time and the second agents' reputation is below ρ_{min}, the first agent defects.

The server computes reputation scores using available feedbacks and using any implemented algorithm. The results of the simulation include: reputation scores of individual agents and the total payoffs (from all auctions) of every agent. The payoffs are affected by the way the reputation system works. For example, if agents post very few feedbacks, reputation scores will be generally random, and the payoffs of good agents would drop. The simulator allows to check whether the implemented reputation algorithm is effective. To verify the concept of implicit feedbacks, we simulate the behavior of a simple reputation algorithm that uses implicit feedbacks.

Consider a user u with the history of n auctions. Let us assume that only $m \leq n$ of these auctions have a feedback. Out of these m feedbacks m^+ are positive feedbacks, while $m^- = m - m^+$ are all other feedbacks. Thus, $m^+ \leq m \leq n$. The reputation ρ_u of the user u will be calculated as follows

$$\rho_u = \frac{m^+}{\alpha * m^- + m}$$

where $0 \leq \alpha \leq 1$. Thus, if $\alpha = 0$, the above reputation score becomes a simple ratio of the number of positive feedbacks received by the user u. In the case when the user u has had no auctions, the above formula is undefined. In such case we set the reputation ρ_u to an initial value, ρ_0. To be precise, in our simulations we use a slightly more complex version of the above algorithm. Since agents in the simulator choose whom they want to interact with on the basis of reputation scores, it is necessary to avoid that the reputation would drop suddenly to a low level. This can happen in the initial phase of the simulation, when the reputation score has not yet stabilized (initially, a single negative feedback could decrease the initial reputation by a large degree). Therefore, we use a moving average to smooth reputation changes. The smoothed reputation $\rho_u^{ma}(t) = 0.5\rho_u^{ma}(t-1) + \rho_u(t)$, where t is time, and $\rho_u^{ma}(0) = \rho_0$ - the smoothed reputation is initialized by the initial reputation value. Note that over time, the estimate converges to the formula for ρ_u (since the impact of the initial reputation decreases exponentially).

4.2 Evalutation by Simulation

We have tested the algorithm described above using the following simulation scenario. First, we have divided all 300 agents into two sets, the *good agents* and the *bad agents*. Good agents were 66% of all agents, the remaining agents were bad agents. A good agent used the Tit-for-Tat strategy with the reputation threshold of $\rho_{min} = 0.5$. A bad agent used a strategy of random cheating with probability $c = 0.6$. All agents had the same behavior with respect to feedbacks. This behavior was a further parameter of the simulation scenarios. We used two posting behaviors: *perfect feedbacks*, where all agents always posted feedback truthfully, and *poor feedbacks*, where if the feedback was positive, an agent would post it with probability $r^+ = 0.66$, and if the feedback was not positive, an agent would post it with probability $r^- = 0.05$. All feedbacks were always true, if they were sent. The parameters of the poor feedbacks were derived from the analysis of traces obtained from the real-world data. In all simulations, 40 000 auctions were simulated between the agents.

Together, there are three significant simulation scenarios: perfect reports with reputation calculated using a simple ratio of positive feedbacks (a reputation algorithm like described in the previous section, only with $\alpha = 0$); poor reports with a simple ratio; and poor reports with the reputation algorithm that uses implicit feedbacks, with different settings for α.

All experiments were conducted using the Monte-Carlo method. We present average results from 10 simulation runs, together with 95% confidence intervals of results. The outcomes of the experiments were the payoffs of every agent. We evaluate the effectiveness of a reputation system using the following criteria: the average payoff of a good agent, the average payoff of a bad agent, and the Gini coefficient of the payoffs of the good agents. The last criterion was introduced as a way of evaluating the effectiveness of the reputation system in providing fairness of the treatment of good agents.

The results of the simulations are summarized in Table 2. The first row in the table corresponds to the perfect feedback scenario, where all agents always post truthful feedbacks, and reputation is calculated using a simple ratio of positive feedbacks. In this idealized scenario, the average payoff of good agents is the highest, at 101.76 (the values of payoffs for a single auction were derived from the payoff table of the Prisoner's dilemma game). The average payoff of a bad agent is much lower, at 22.66. This indicates that the reputation mechanism

Table 2. Impact of implicit feedback reputation algorithm on agent payoffs and justice

Scenario	AVGP+[a]	AVGP-[b]	GC+[c]	GCI[d]	AGPCI[e]
Perfect reports	101.76	22.66	0.70	0.63–0.76	100–103,5
Poor reports, $\alpha = 0$	96.45	54.20	0.51	0.45–0.58	93.6–99.3
Poor reports, $\alpha = 0.05$	99.08	23.03	0.75	0.66–0.85	96.1–102
Poor reports, $\alpha = 0.1$	100.40	23.52	0.67	0.58–0.76	98.8–101.9
Poor reports, $\alpha = 0.2$	100.74	22.64	0.74	0.66–0.82	98–103.4

[a] Average payoff of a good agent.
[b] Average payoff of a bad agent.
[c] Gini coefficient of good agents.
[d] Gini confidence interval (95%).
[e] Average good payoff confidence interval (95%).

is working because cheating agents get punished by lower reputation. In these simulations, the final reputation value of bad agents was almost always 0. The Gini coefficient at about 0.7 will be treated as a reference level for further experiments and values of the Gini coefficient above this level will be considered unacceptable. The 95% confidence intervals for both the Gini coefficient and the average payoff are quite narrow. The second row of the Table 2 shows the results of the second simulation scenario. In this scenario, agents provided feedbacks realistically, and the effect of this is an increase in the payoff of bad agents by almost 150%. In many simulations, bad agents managed to keep a high reputation value, leading the good agents to trust them. This enabled bad agents to cheat more good agents. As a result, the average payoff of good agents also decreased significantly. This decrease is also visible in the confidence interval of payoffs of good agents.

Further rows of the table show the impact of using implicit feedbacks. The rows correspond to using the reputation algorithm described in the previous section with different values of α. For all considered values of α, the payoffs of bad agents dropped sharply, almost to the level achieved when agents reported perfectly. This is the main argument for using implicit feedback: as our simulations indicate, the use of implicit feedbacks is efficient in preventing cheating. The payoffs of good agents also increased to a varying degree, but for all values of α, the average payoff of a good agent was higher than when a simple ratio of positive feedbacks was used as the reputation algorithm. On the basis of the performed experiments, it seems that the value of $\alpha = 0.1$ gave the best results. For $\alpha = 0.2$, the average payoffs of the good agents were higher, and the average payoffs of bad agents were lower than for $\alpha = 0.1$. However, the average Gini coefficient was also higher. The reason for this may be that in the simulations, good agents sent positive feedbacks randomly with a probability of 66%. It was possible that a good agent would repeatedly get no positive feedback for her cooperation with another good agent. This could result in decreasing the reputation of the good agent, especially for higher values of α. The poor performance of $\alpha = 0.05$ can be explained by the fact that with such a low setting of α, the reputation of bad agents did not decrease quickly enough. While our simulations do not allow to choose the value of α that would be applicable in a real-world scenario, they are sufficient to indicate that there should exist an optimal value of α that is neither too high nor too low.

4.3 Evaluation by Mining

We conduct the experiments on a large body of real-world data acquired from `www.allegro.pl`, the leading provider of online auctions in Poland. The examined dataset consists of 656 376 auctions collected over the period of six months. The reputation of users is determined based on 890 876 mutual feedbacks provided by users. The dataset has been created by gathering all auctions of a seed set of 10 000 buyers during a fixed period of six months.

Figure 3 presents the influence of implicit feedbacks discovered using the majority voting strategy on the reputation score. Users are ordered according to

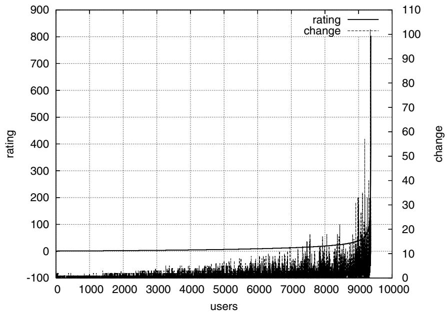

Fig. 3. Influence of missing feedbacks (majority voting strategy)

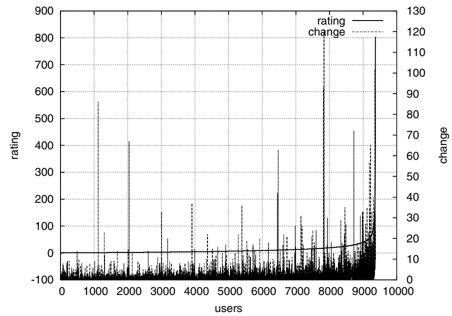

Fig. 4. Influence of missing feedbacks (cosine strategy

their rating computed traditionally and reported on the first y-axis. The second y-axis represents the change in user's reputation if implicit feedbacks are taken into consideration. We give an implicit feedback the weight of 0.2 of the weight of a negative feedback. On average, the reputation of users drops by 15% when implicit feedbacks are included in reputation scoring. For many users the change in reputation score is negligible, but there are users for whom the change is significant. We believe that traditional reputation systems grossly overestimate the reputation of these buyers and only incorporating the implicit feedback reveals their poor performance. The results of a similar experiment are depicted in Figure 4. This time we use the cosine strategy with the Ochini coefficient threshold $\beta = 0.4$. As can be easily noticed, the cosine strategy identifies more missing feedbacks as implicit feedbacks, thus affecting the reputation scoring stronger than the majority voting strategy. On average, the reputation of users drops by 17% when implicit feedbacks are included in reputation scoring. Of course, the impact factor depends on the value of the Ochini coefficient threshold. It remains to be seen which value of the threshold produces the most accurate and credible identification of implicit feedbacks. The results of both experiments affirm the practical usability and importance of using implicit feedbacks.

5 Conclusions

In this paper we have introduced the notion of the implicit feedback. To the best of authors' knowledge this is the first proposal to mine online auction data in search of unvoiced assessments of other user performance. We have been able to show that the use of implicit feedbacks in a reputation system can be effective. A simple reputation algorithm that used implicit feedbacks outperformed the reputation algorithm used by eBay. Using implicit feedbacks prevents the overestimation of reputation of dishonest auction participants, because negative opinion about their performance, otherwise concealed by intimidated users, can be used in reputation scoring. In this paper we have presented our initial findings on the impact of using implicit feedback with a simple reputation system. We plan to extend our experiments on more complex reputation algorithms [8].

References

1. Robert M. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.
2. Sulin Ba, Andrew B. Whinston, and Han Zhang. Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems*, 35:273–286, 2003.
3. Piotr Gmytrasiewicz and E. Durfee. Toward a theory of honesty and trust among communicating autonomous agents. *Group Decision and Negotiation*, 2:237–258, 1993.
4. Michael M. Kostreva, Wlodzimierz Ogryczak, and Adam Wierzbicki. Equitable aggregations and multiple criteria analysis. *European Journal of Operational Research*, 158(2):362–377, October 2004.
5. R. A. Malaga. Web-based reputation management systems: Problems and suggested solutions. *Electronic Commerce Research*, 4(1), 2001.
6. Albert W. Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.
7. Mikhail I. Melnik and James Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. *The Journal of Industrial Economics*, L(3), September 2002.
8. Mikołaj Morzy. New algorithms for mining the reputation of participants of online auctions. In *WINE 2005, 1st Workshop on Internet and Network Economics, 15-17 December 2005, Hong Kong*. Springer Verlag, 2005.
9. Lik Mui. *Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks*. PhD thesis, Massachusetts Institute of Technology, 2003.
10. Paul Resnick and Richard Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *Advances in Applied Microeconomics*, 11, 2002.
11. Paul Resnick, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara. Reputation systems. *Communications of the ACM*, 43(12), 2000.
12. Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on ebay: a controlled experiment. Technical report, School of Information, University of Michigan, 2004.