On Assigning Implicit Reputation Scores in an Online Labor Marketplace

Maria Daltayanni
UC Santa Cruz
mariadal@cs.ucsc.edu

Luca de Alfaro
UC Santa Cruz
luca@cs.ucsc.edu

Panagiotis Papadimitriou
oDesk Research
papadimitriou@odesk.com

Panayiotis Tsaparas
University of Ioannina
tsap@cs.uoi.gr

ABSTRACT

In online labor marketplaces employers post job openings and receive applications by workers interested in them. The employers decide which applicant to hire and then they work with the selected worker to accomplish the job requirements. At the end of the contract, an employer can provide his worker with some rating that becomes visible in the online worker profile and can guide future hiring decisions of other employers. In this paper, we discuss some of the shortcomings of the existing reputation system and we propose a new reputation mechanism that combines employer implicit feedback signals in a link-analysis-based approach. The new system addresses the shortcomings of the existing one while yielding similar or better signal for the worker quality.

Categories and Subject Descriptors

H.4.4.3.4 [**Information Systems**]: World Wide Web—*Web applications, Crowdsourcing, Reputation Systems*

General Terms

Algorithms, Design, Experimentation, Performance

Keywords

Link Analysis, Reputation Systems, Crowdsourcing, Labor Marketplace

1. INTRODUCTION

In online labor marketplaces, such as oDesk.com, Elance and Freelancer.com, two parties are involved; employers and workers. Employers post job openings and candidate workers apply to them, based on their qualifications, skills and interests. The employers review the applicants' online resumes, and interview few applicants to take hiring decisions. The worker reputation, i.e., the ratings that the worker has received in his past jobs in the platform, is one of the most important considerations for the employer hiring, since it reveals how other employers evaluate the worker true ability in real job scenarios. Although the reputation information is a useful signal, it is usually very *sparse*, since a worker needs to apply, get hired and complete few jobs before he obtains a representative reputation score. The reputation scores are also *skewed* towards high ratings [1], because employers care about the impact of their feedbacks on the workers' future opportunities for jobs in the marketplace. The skewed distribution of ratings make them less helpful in identifying very competent workers.

To address the limitations of existing reputation systems in labor marketplaces, we present *WorkerRank*, a new reputation system that leverages employers' implicit judgements at the application evaluation moment, rather than the employer's explicit feedback at the job completion moment. Although the implicit judgments are more noisy than the explicit ones, they are more broadly available, since the number of applications is usually one to two orders of magnitudes higher than the number of hires. Moreover, the implicit actions of the employers are not revealed and consequently the employers do not bias their judgments towards high ratings (as happens when they aim to avoid the negative impact on the workers). As a result, the obtained ratings are not skewed. To deal with the noise of implicit judgments we present various weighting schemes (Section 2) that we evaluate on a real-world dataset from oDesk (Section 3). Our results show that the new reputation system not only provides information for far more workers in the marketplace, but it also serves as a better discriminatory signal for hiring decisions.

2. WorkerRank MODEL

In this section we present the basic elements of WorkerRank. We represent the marketplace data with a symmetric directed bipartite graph $G = (U, V, E)$ (see figure 1); U is the set of jobs posted by employers within a specific time period; V is the set of workers who applied to the posted jobs. An edge $(v, u) \in E$ represents the application of the worker $v \in V$ to job $u \in U$. The edge $(u, v) \in E$ represents how successful the application was, based on the employer's decision. We label edges with employer responses: {offer, interview, shortlist, ignore, hide, reject}.

We build the reputation system WorkerRank that is based on the
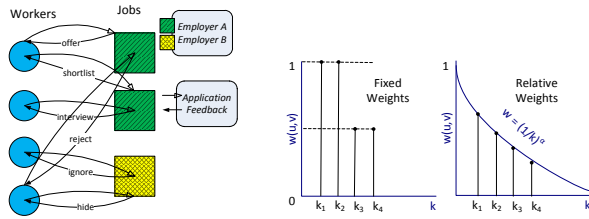
**Figure 1: Left: Bipartite graph between workers and jobs posted by employers, Right: Weight Graphs**

Figure 2: Time needed to obtain reputation signal for new workers

HITS algorithm [2]. Each node is assigned a score based on the cummulative scores of the incoming edge source nodes. Each worker node $v \in V$ is assigned a *reputation* score $r(v) \in \mathbb{R}$ and each job node $u \in U$ is assigned a score $b(u) \in \mathbb{R}$. In the extended version of this paper, we describe weighing schemes on the edges (figure 1), that aim at capturing the amount of reputation transmitted by each edge, and achieve a more precise reputation. Here we only give a brief description of the main contributions of the weighing schemes, and of the metrics, towards producing objective reputation scores.

In the first approach we assign *fixed* weights to edges for the different application success types. Since edges carry the success information, we assume that edge weights are proportional to the application success level for all candidates of a given job. For example, a candidate with successful applications deserves higher credit than candidates whose applications were rejected. Fixed weights overlook the fact that the data are structured in job-wise fashion. In fact, application success of an applicant is not independent from the application success of the remaining candidates at a particular job, since his hire probability is affected by their performance. Hence in the second approach we allow for *relative ranks* of candidates compared to the ranks of other candidates, when ranked by application success in that job. Finally, a job may have twice as many applicants as another job; gaining an offer would then be more competitive in the first and the recipient should receive higher credit. In the third approach we multiply the above weighing schemes by $\frac{n-n_l}{n}$, where n is the total number of applicants to a job and n_l is the total number of applicants to the job with label l. In the context of job-wise structured data, competitiveness serves as a normalization factor along weights in different jobs of the marketplace.

3. EXPERIMENTAL RESULTS

We performed experiments to evaluate WorkerRank and compare the weighting schemes we briefly discussed in Section 2. For our experiments we used a sample of real-world application data and reputation scores provided by oDesk. The dataset spans the time period of 53 weeks between March 2012 through March 2013 and it contains approximately 17M applications submitted by 0.5M workers to 0.8M openings posted by 0.2M employers. In the following three paragraphs we compare WorkerRank with the existing reputation scheme in terms of (a) the sparsity of the signal in the marketplace, (b) the time needed to obtain a signal for new workers, and (c) discriminatory power for hiring decisions.

We run the WorkerRank algorithm over the applications of the first 52 weeks of our dataset. During this time period we also keep track of the ratings that the workers receive after the end of accomplished jobs. Then we report the number of applications of the 53th
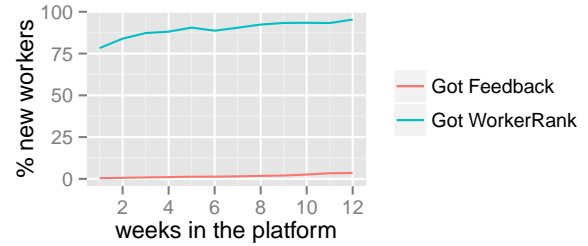
week for which there is a WorkerRank score versus the applications for which there is an employer feedback score. Our results show that out of $88,294$ applications in the 53th week, we have WorkerRank scores for $79,083$ (89.6%), while we have feedback scores only for $52,471$ (59.4%). The increase in the marketplace application coverage is 50.7%. Since the online marketplaces grow fast, the identification of new competent workers is very significant for their healthy development. For all workers who joined the oDesk platform during the last 12 weeks of our study period, we calculate the percentage of workers for which we obtain reputation signals within X weeks. X is varying from 1 to 12 weeks. As presented in Figure 2, the WorkerRank scores are available for more than 75% of the new workers within one week of their joining the platform and the pertentage ratio grows to 95% after 12 weeks. On the contrary, there are less than 1% of new workers who received feedback at the end of their first week at platform and this percentage does not exceed 5% at the end of the 12-week period.

Finally, to evaluate the quality of the WorkerRank scores, we do compare it with the existing reputation scores as signals for taking hiring decisions. We use the data of the first 52 weeks of our dataset to calculate the WorkerRank scores for all of the WorkerRank variations presented in Section 2. We then use each score as a unique predictor for the hiring outcome of the applications submitted during the 53th week. We compare the various models by ranking the applicants in every opening according to their reputation score and calculate the Mean Average Precision (MAP) across all openings. We present the results in Figure 3 where we see that the performance of all WorkerRank variations is around 20%. Such a score means that the expected position of the applicant to be hired is the 5th, while the average number of applicants per opening is around 20. Note also that the WorkerRank variations achieve similar performance to the reputation system and the 'selective' weighting scheme even improves it. These results show that the quality of the suggested reputation system is similar to or better than the existing baseline.

4. CONCLUSION

The results of our experiments show that WorkerRank improves rank prediction accuracy compared to baseline approaches. What is more, WorkerRank solves the basic problems encountered in industrial reputation systems (unreliable employer ratings, coverage, cold start). Our future work includes research on weighting schemas and modeling implicit actions on the marketplace website.

5. REFERENCES
[1] N. Hu, J. Zhang, and P. A. Pavlou. Overcoming the j-shaped distribution of product reviews. *Commun. ACM*,
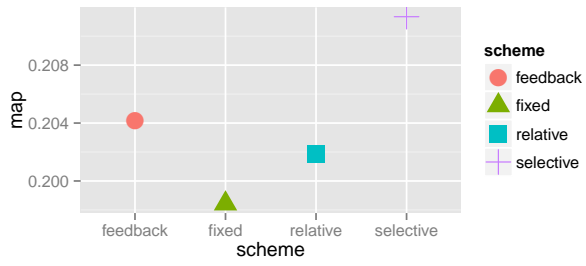
Figure 3: MAP for predicting the hiring outcome

52(10):144–147, Oct. 2009.

[2] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, Sept. 1999.