Dynamic Recommendations for Sequential Hiring Decisions in Online Labor Markets

Marios Kokkodis
Carroll School of Management
Boston College
kokkodis@bc.edu

ABSTRACT

Online labor markets facilitate transactions between employers and a diverse set of independent contractors around the globe. When making hiring decisions in these markets, employers have to assess a large and heterogeneous population of contractors. Because many of the contractors' characteristics are latent, employers often make risky decisions that end up in negative outcomes. In this work, we address this issue by proposing a framework for recommending contractors who are likely to get hired and successfully complete the task at hand. We start our analysis by acknowledging that employers' hiring behavior dynamically evolves with time; Employers learn to choose contractors according to the outcomes of their previously completed tasks. To capture this dynamic evolution, we propose a structured Hidden Markov Model that explicitly models task outcomes through the employers' evolution. We build and evaluate the proposed framework on a dataset of real online hiring decisions. We then compare our approach with a set of previously proposed static algorithms and we show that our proposed framework provides up to 24% improved recommendations. We conclude by discussing the positive impact that such better recommendations of candidates can have on employers, contractors, and the market itself.

CCS CONCEPTS

• **Information systems → Data analytics**;

KEYWORDS

Online labor markets, Dynamic hiring decisions, HMM

ACM Reference Format:
Marios Kokkodis. 2018. Dynamic Recommendations for Sequential Hiring Decisions in Online Labor Markets. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3219881

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219881

1 INTRODUCTION

Online labor markets such as Peopleperhour.com and Freelancer.com connect employers with contractors around the globe to accomplish a diverse set tasks. These tasks span multiple categories, including web development, graphic design, accounting, sales and marketing, etc. On par with other online platforms (e.g., Amazon.com [22]), online labor marketplaces have experienced an accelerated growth in the past decade.[1] A typical transaction in these markets starts with an employer posting a job description. Online contractors who are looking for opportunities decide whether or not to bid for the job. From the pool of contractors who finally bid, the employer chooses who to hire. Once hired, the contractors complete the task and receive the pre-arranged payments.

To identify the best candidate for the opening at hand, employers assess a series of observed and latent characteristics of the available contractors. The observed characteristics are usually the contractors' skills, work histories, and certifications. The latent characteristics are the contractors' true knowledge and abilities [13]. The existence of latent characteristics, the heterogeneity that appears in the observed ones [23], and the interactions between the two create an uncertain environment of information asymmetry [3].

To minimize this information asymmetry and help employers make better hiring decisions, online labor markets have developed reputation systems. Specifically, contractors get rated for the tasks they complete and these ratings become part of their online resumes. Employers then consider the contractors' past performance to make better-informed hiring decisions. In practice however, reputation scores in these markets are inflated [18, 20]. This inflation renders feedback scores uninformative and sometimes even misleading. Because of that, online labor markets have introduced various recommendation frameworks. These frameworks rank contractors according to their likelihood of getting hired, and as a result they provide a refined, curated list of recommendations from which employers can make better-informed hiring decisions [27].

Even though these ranking mechanisms are a step in the right direction, they suffer from two major shortcomings. First, they do not account for the task outcomes. Specifically, online labor markets implement a bilateral feedback mechanism, where both the contractor and the employer rate the outcome of their interaction upon the completion of a task. Figure 1 shows that these outcomes are highly heterogenous. As a result, any recommendation algorithm that ignores these outcomes (such as those presented by Kokkodis *et al.* [27]) will include recommendations that reinforce suboptimal behaviors.

The second shortcoming of the current recommendation algorithms is that they do not take into consideration the correlations

[1] http://www.statista.com/topics/871/online-shopping

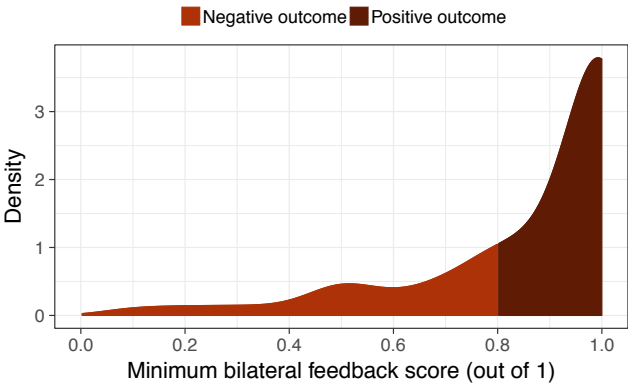

Figure 1: The distribution of hiring outcomes. We assume that outcomes with minimum bilateral feedback scores greater than 0.8 are positive, while outcomes with scores lower than 0.8 are negative.

between sequential hiring decisions of the same employers. In every online labor market, a significant portion of the revenue comes from employers who hire repeatedly multiple contractors for various tasks. Previous algorithms treat hiring decisions of the same employers as independent [27]. However, it is natural to assume that employers actually learn through the outcomes of previously completed tasks, and as a result evolve their hiring behaviors.

In this work, we focus on addressing these two shortcomings. First, we train models that rank candidates according to their likelihood of getting hired and completing the task successfully Second, we acknowledge that employers' preferences and hiring processes dynamically evolve through repeated hires. To capture this evolution, we propose a structured Hidden Markov Model, where we allow (1) the outcomes of previous transactions to define the transition probabilities between different states, and (2) the job-opening-contractor characteristics to predict the outcome of a job application.

We build and evaluate the proposed framework on a set of hiring decisions from an online labor market. We compare the performance of our approach with the ranking mechanisms described in Kokkodis *et al.* [27] and show that the structured HMM performs significantly better. In addition, we show that the true benefit of our framework is in repeated hiring decisions of the same employers, where it curates recommendations that are up to 24% better than the resulting recommendations of the static ranking mechanisms.

Our work is the first to propose a dynamic framework that explicitly models the evolving and heterogeneous nature of employers' hiring behaviors. We show evidence that our approach provides better recommendations than the previously proposed static ranking implementations. Better recommendations can benefit online labor markets in three major ways. First, employers can make better-informed and faster decisions based on the suggested applicants. Second, contractors can save time by not applying to openings that have very low hiring probabilities. Third, because the outcomes of the recommendation-based transactions should be on expectation positive (since this is what our approach optimizes), both the employers and the contractors will have the incentives to keep participating. Simply put, our proposed framework has the potential to minimize the employers' search cost and increase the market's transaction volume through the overall increase in satisfaction of the contractors and the employers.

2 BACKGROUND

The recent emergence and growth of online labor markets [1] has triggered research on various aspects of hiring decisions. Because of the global reach of these markets, previous works have focused on the importance of location in hiring decisions. In particular, researchers have shown that (1) employers show a strong preference towards contractors from their own country [11, 14, 17, 31], (2) contractors from less developed countries are disadvantaged compared to those from developed countries in terms of their likelihood of being hired [2, 17, 32], and (3) contractors' reputation attenuates the negative effects of location [17]. In this work, we use a location feature (point-wise mutual information–section 4.2) that captures the association between the contractor-employer locations and the contractor's hiring probability.

A significant portion of previous related research focuses on the importance of reputation and its role in hiring decisions. Researchers have found that (1) reputation is an important predictor of success only for fixed-price contracts [30], (2) employers place a significant weight on contractors' reputation when making hiring decisions [39], (3) employers trade-off reputation and price and are willing to accept higher bids posted by more reputable contractors [33], (4) reputation has a significant impact on the contractors' wage [6] and it transfers among certain task categories [24, 25], and (5) both hiring contractors and providing detailed evaluations substantially improve contractors' subsequent employment outcomes [35]. For the proposed framework, we use reputation signals (i.e., the accumulated feedback score, the total number of completed jobs, and the total number of completed work-hours) as predictors of the likelihood of a contractor to get hired and complete the tasks successfully.

A main focus of our study is to explicitly model the heterogenous and evolving nature of employers' hiring behavior. We know from previous works that when employers signal their price and quality preferences, the employer-contractor matchings significantly improve [19]. More recently, we learned that employers show a positive hiring bias in favor of female contractors [9]. Furthermore, after a transaction with a positive outcome, employers show a positive bias to the contractor's country [32]. These studies capture some of the behavioral dimensions of employers. However, they do not explicitly model the employers' evolving process through consecutive hiring outcomes. Even though our work focuses on presentig a recommendation framework to guide hiring decisions, we extend these works by providing a modeling approach that could be potentially adjusted to study how hiring preferences evolve through consecutive completed tasks.

In a series of other related works on hiring decisions, researchers have found that the presence of agencies signals increased quality for inexperienced contractors [38], and that high-value projects attract significantly more bids [12, 37]. Finally, Goes and Lin [16] found that certification status can negatively impact contractors'

likelihood of getting hired. In Section 4.2 we use the results of these studies to conceptualize the introduction of certain employer-contractor-opening features in our predictive framework.

This work directly extends the previous work of Kokkodis *et al.* [27]. In their paper the authors proposed a series of recommendation algorithms that reinforce the hiring behaviors of employers in an online labor marker. The proposed models however did not account for (1) the hiring outcome and (2) the evolution of employers through consecutive hiring decisions. We extend their work by proposing a framework that explicitly addresses these two points, and we show that such a framework results in better recommendations. We discuss this framework next.

3 PROBLEM FORMULATION AND METHODOLOGY

An online labor market transaction starts with an employer creating an opening by posting a job description. The job description reveals a series of characteristics that the employer is looking for, such as the required set of skills and experience. The job description and the related characteristics are job-specific and they are fully observed by the interested contractors. Some of these contractors who are on the look-out for opportunities and see a fit with the opening submit their applications. For each application that arrives, the employer makes a hiring decision based (1) on the employer's characteristics $\boldsymbol{X}_i$, (2) the applicant characteristics ($\boldsymbol{X}_a$), and (3) the opening characteristics ($\boldsymbol{X}_o$). For simplicity, we merge the three vectors into $\boldsymbol{X} = [\boldsymbol{X}_i, \boldsymbol{X}_o, \boldsymbol{X}_a]'$.

Problem definition: *Given the set of available contractors, the employer, and the task at hand, our goal is to recommend the contractors who are more likely to get hired and complete the task successfully.*

HMM formulation: The employer's hiring decision process is latent (unobserved). From this process we only observe the outcome of each job application. In particular, for each applicant the employer must first choose whether or not to hire. If the employer chooses to not hire the applicant, we observe a "no hire" outcome. If the employer chooses to hire the applicant, we observe the outcome of the transaction after the completion of the task. This outcome is defined by the minimum of the bilateral feedback scores that employers and contractors assign to each other upon the completion of the task. For simplicity, we assume a positive outcome when the minimum assigned score is greater than 0.8, and a negative outcome otherwise (see Figure 1).[2] Formally, we observe the following set of possible outcomes ($\mathcal{Y}$):

$$\mathcal{Y} = \{\text{no hire, negative, positive}\} \ . \tag{1}$$

Employers' decisions evolve dynamically with time. Specifically, we expect employers to learn and optimize their hiring evaluation process based on the outcomes of their previous transactions on the platform. We formulate this evolution as a structured Hidden Markov Model (HMM).[3] At any point in time t, employers operate from a latent state, which determines their propensity to make a hiring decision with an outcome $Y_t \in \mathcal{Y}$. Depending on the outcomes of these hiring decisions, employers stochastically transition to new latent states. For our HMM, we assume a set of K latent states ($\mathcal{S}$):

$$\mathcal{S} = \{s_1, s_2, \ldots, s_K\} \ . \tag{2}$$

For the rest of the paper we assume that the unit of time (t) is a new hiring decision.

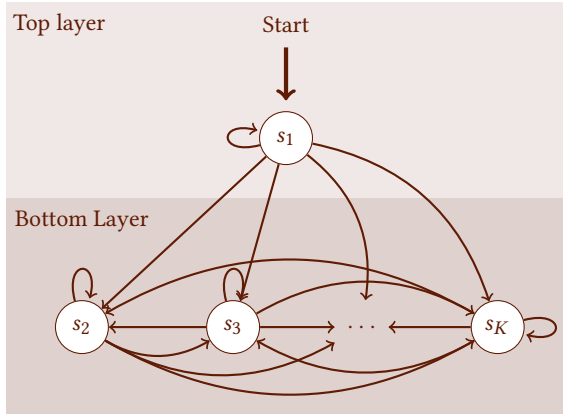

Figure 2: The structure of the two-layered HMM. In the top layer, there is only the initial state s_1, where we cluster all employers when they first join the platform. Once the employers hire contractors and complete transactions on the platform, they emit positive or negative outcomes and they transition to an appropriate state in the bottom layer. In the bottom layer, the employers are not allowed to return to the top layer, but instead, they are allowed to stochastically transition in the $K-1$ available states.

HMM structure: Every new employer who joins the platform has certain preferences and biases when choosing contractors to hire. As the employer rejects some applicants and hire others, we observe signals of the employer's hiring process. To capture this behavior in our HMM, we impose a two-layered structure: In the first layer, we assume only an initial latent state s_1, where we land all new employers. We use this state to make an initial estimate of the employers' hiring processes.[4] Once the employers complete their first transaction and emit a positive or a negative outcome, they stochastically transition to one of the $K-1$ states of the second layer. We show this structure in Figure 2, where we distinguish the two layers and derive the possible transitions from each state.

To completely define an HMM, we need (1) a vector of initial state probabilities $\boldsymbol{\pi}$, (2) a transition matrix T that holds the transition probabilities between states, and (3) an emission matrix E that describes the state-specific probability distributions across the set of observations $\mathcal{Y}$. Since we assume that every new employer lands in state s_1, we define the initial probability vector for our HMM as follows:

$$\boldsymbol{\pi} = [1, 0, 0, \ldots, 0]' \ . \tag{3}$$

[2]Extensions to more granular outcomes are conceptually trivial, but could result in very sparse observations.

[3]HMMs have been used in modeling diverse sequential environments, including user contribution in online communities [26] and online advertising [15].

[4]The alternative here would have been to land employers stochastically to one of the second layer states. This would increase the noise in our predictions, and as a result it would hurt the performance of our approach.

Even though an employer emits observations $Y_t \in \mathcal{Y}$ for every received application, we allow employers to transition to a new state only after a positive or a negative outcome. The intuition is that employers learn and adjust their hiring behaviors only through completed transactions. Specifically, we assume that the history of positive and negative outcomes of each employer directly affects the employer's transition to a new state. Hence, we define the vector $\boldsymbol{Z}_t = [\xi_t, \zeta_t]'$, where ξ_t is the number of previous positive outcomes at time t and ζ_t is the number of previous negative outcomes at time t.

Let us denote the transition probability of a given employer to move from state s_k to state s_l at time t as follows:

$$\lambda^{s_k s_l}_{\boldsymbol{\gamma}_{kl} \boldsymbol{Z}_{t-1}} = \Pr(S_t = s_l | S_{t-1} = s_k; \boldsymbol{\gamma}_{kl}, \boldsymbol{Z}_{t-1}) \ . \tag{4}$$

In the previous Equation, $\boldsymbol{\gamma}_{kl}$ is the vector of coefficients of state s_k that define the weights of ξ_{t-1}, ζ_{t-1} in estimating the transition probability to state s_l. We estimate this probability through the following activation function:

$$\lambda^{s_k s_l}_{\boldsymbol{\gamma}_{kl} \boldsymbol{Z}_{t-1}} = \begin{cases} 0, & \text{if } \xi_{t-1} + \zeta_{t-1} = 0, s_l \neq s_k \\ 1, & \text{if } \xi_{t-1} + \zeta_{t-1} = 0, s_l = s_k \\ \text{softmax}(\boldsymbol{\gamma}_{kl} \boldsymbol{Z}_{t-1}), & \text{if } \xi_{t-1} + \zeta_{t-1} > 0 \ . \end{cases} \tag{5}$$

Because of the structure that we impose to our HMM (figure 2), the transition matrix is not completely filled with elements of Equation 5. Specifically, at time t, the transition matrix has the following form:

$$T(\boldsymbol{\gamma}, \boldsymbol{Z}_{t-1}) = \begin{bmatrix} \lambda^{s_1 s_1}_{\boldsymbol{\gamma}_{11} \boldsymbol{Z}_{t-1}} & \lambda^{s_1 s_2}_{\boldsymbol{\gamma}_{12} \boldsymbol{Z}_{t-1}} & \cdots & \lambda^{s_1 s_K}_{\boldsymbol{\gamma}_{1K} \boldsymbol{Z}_{t-1}} \\ 0 & \lambda^{s_2 s_2}_{\boldsymbol{\gamma}_{22} \boldsymbol{Z}_{t-1}} & \cdots & \lambda^{s_2 s_K}_{\boldsymbol{\gamma}_{2K} \boldsymbol{Z}_{t-1}} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \lambda^{s_K s_2}_{\boldsymbol{\gamma}_{K2} \boldsymbol{Z}_{t-1}} & \vdots & \lambda^{s_K s_K}_{\boldsymbol{\gamma}_{KK} \boldsymbol{Z}_{t-1}} \end{bmatrix} \tag{6}$$

In the previous, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_{11}, \boldsymbol{\gamma}_{12}, \ldots, \boldsymbol{\gamma}_{KK}]'$.

To completely define the HMM we need to derive its emission matrix. This matrix, consists of elements that describe the conditional probabilities of observations given the current state of the employer. We assume that these emission probabilities follow a multinomial distribution across our set of available observations, $\mathcal{Y}$. We further assume that the emission probabilities are directly affected by the individual features of the employer, the applicant, and the opening, i.e., the vector $\mathbf{X}_t$ at time t. Formally, we compute the emission probabilities as follows:

$$\Pr(Y_t = y | S_t = s_k; \boldsymbol{\beta}_{k,y}, \boldsymbol{X}_t) = \text{softmax}(\boldsymbol{\beta}_{k,y} \mathbf{X}_t) \ , \tag{7}$$

where $y \in \mathcal{Y}, s_k \in \mathcal{S}$, and $\boldsymbol{\beta}_{k,y}$ is the vector of coefficients for estimating the probability to observe outcome y when being in state s_k and having an input vector of features $\boldsymbol{X}_t$. Finally, we define the complete vector $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_{1,\text{no hire}}, \boldsymbol{\beta}_{1,\text{negative}}, \boldsymbol{\beta}_{1,\text{positive}}, \ldots, \boldsymbol{\beta}_{K,\text{positive}}]' . \tag{8}$$

HMM identification: Given the structure of this HMM, we need to estimate the parameter vectors $\boldsymbol{\beta}, \boldsymbol{\gamma}$. To do so, we maximize the conditional probability of the set of observations given the HMM. Let us assume that we have the following sequence of M observations for a given employer i:

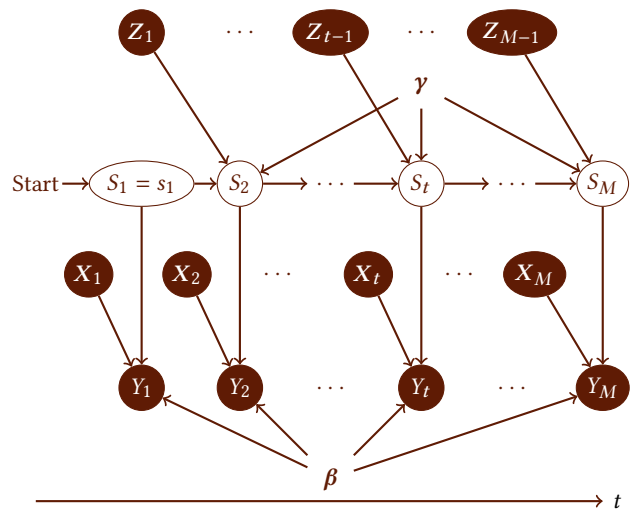

Figure 3: The structure of the latent state sequence S_i, the observed sequence of outcomes Y_i, the parameter vectors $\boldsymbol{\beta}, \boldsymbol{\gamma}$, and the sequences of input vectors $X_{1:M}, Z_{1:M-1}$ for a given employer i. For better readability we drop the employer subscript i. As with traditional probabilistic graphical models, we show the latent states in clear ellipses, while everything we observe we show in shaded ones.

$$\boldsymbol{Y}_i = Y_{i1}, Y_{i2}, \ldots, Y_{iM} \ , \tag{9}$$

where $Y_{im} \in \mathcal{Y}, m \in \{1, 2, ..., M\}$. These observations correspond to a sequence of input vectors:

$$\boldsymbol{X}_{1:M} = \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M \ . \tag{10}$$

Furthermore, let us assume that $\boldsymbol{Y}_i$ is the result of a sequence of latent states, $\boldsymbol{S}_i$:

$$\boldsymbol{S}_i = S_{i1}, S_{i2}, \ldots, S_{iM} \ , \tag{11}$$

where $S_{im} \in \mathcal{S}$. This sequence of states is affected by the sequence of outcome vectors $\boldsymbol{Z}_{1:M-1}$:

$$\boldsymbol{Z}_{1:M-1} = \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_{M-1} \ . \tag{12}$$

We show these sequences along with their interactions in Figure 3. Based on the structure of our graph, we get the conditional likelihood of observing $\boldsymbol{Y}_i$:

$$\Pr(\boldsymbol{Y}_i | \boldsymbol{S}_i; \boldsymbol{\beta}, \boldsymbol{X}_{1:M}) = \prod_{t=1}^{M} \Pr(Y_{it} | S_{it}; \boldsymbol{\beta}, \boldsymbol{X}_t) \ , \tag{13}$$

where we estimate the right hand side through Equation 7. From Figure 3 we further get the conditional probability of observing the sequence $\boldsymbol{S}_i$:

$$\Pr(\boldsymbol{S}_i | \boldsymbol{\gamma}, \boldsymbol{Z}_{1:M-1}) = \boldsymbol{\pi}(S_1) \prod_{t=2}^{M} \Pr(S_{it} | S_{it-1}; \boldsymbol{\gamma}, \boldsymbol{Z}_{t-1}) \ , \tag{14}$$

where $\boldsymbol{\pi}(S_1)$ is the the prior probability of being at state S_1. We estimate this Equation through the transition activation function (Equation 5). Since the structure of the HMM imposes that every

new user lands in state s_1 ($\boldsymbol{\pi}(S_1 = s_1) = 1$) we simplify the previous equation to the following:

$$\Pr(\boldsymbol{S}_i|\boldsymbol{\gamma}, \boldsymbol{Z}_{1:M-1}) = \prod_{t=2}^{M} \Pr(S_{it}|S_{it-1}; \boldsymbol{\gamma}, \boldsymbol{Z}_{t-1}) \ . \tag{15}$$

Based on this analysis and the graph in Figure 3 we derive the likelihood of this sequence of observations for employer i as follows:

$$\begin{aligned} l(\boldsymbol{Y}_i; \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \Pr(\boldsymbol{Y}_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{X}_{1:M}, \boldsymbol{Z}_{1:M-1}) \\ &= \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i, \boldsymbol{S}_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{X}_{1:M}, \boldsymbol{Z}_{1:M-1}) \\ &\overset{\text{Figure 3}}{=} \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i|\boldsymbol{S}_i; \boldsymbol{\beta}, \boldsymbol{X}_{1:M}) \Pr(\boldsymbol{S}_i|\boldsymbol{\gamma}, \boldsymbol{Z}_{1:M-1}) \\ &= Pr(Y_{i1}|S_{i1}; \boldsymbol{\beta}, \boldsymbol{X}_1) \\ &\cdot \sum_{\forall \boldsymbol{S}_i} \prod_{t=2}^{M} \Pr(Y_{it}|S_{it}; \boldsymbol{\beta}, \boldsymbol{X}_t) \\ &\cdot \Pr(S_{it}|S_{it-1}; \boldsymbol{\gamma}, \boldsymbol{Z}_{t-1}) \ , \end{aligned} \tag{16}$$

where we used the structure of the HMM to decompose the joint probability of $\Pr(\boldsymbol{Y}_i, \boldsymbol{S}_i|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{X}_{1:M}, \boldsymbol{Z}_{1:M-1})$.

Finally, we get the complete likelihood for a dataset with N employers as follows:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{N} l(\boldsymbol{Y}_i; \boldsymbol{\beta}, \boldsymbol{\gamma}) \ . \tag{17}$$

Once we have the complete likelihood, we can estimate the parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ that maximize it.[5] We do this numerically through the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [8].

4 EXPERIMENTAL SETUP AND RESULTS

Before we analyze the performance of our approach (section 4.4), we need to present (1) the dataset we used, (2) the construction of the feature vector $\boldsymbol{X}$ and (3) our setup.

4.1 Data

To build and evaluate our framework we use a dataset of hiring decisions from an online labor market. This market provides a set of tools for employers who intend to hire and manage remote contractors. We use a set of 392,411 job applications for 17,304 openings that led to 25,297 completed tasks. These tasks span across a series of categories, including web and graphic design, sales and marketing, administrative support, etc. Furthermore, the dat includes both fixed price tasks (i.e., tasks for which the compensation is agreed upon at the signing of the contract) and hourly priced tasks (i.e., the contractors get compensated per hour of work).

In our dataset we observe a total of 367,114 “no hire" outcomes, 10,776 negative and 14,521 positive ones. Figure 4 shows the distributions of the completed tasks and hiring decisions per employer in our dataset. Recall that the completed tasks are the observations that trigger transitions to new states (see Equation 5), while the hiring decisions per employer represent the lengths of sequences that we train our HMM on.

[5]In practice we minimize the negative log-likelihood.

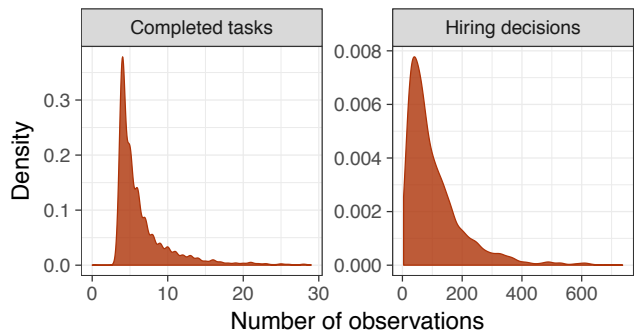

Figure 4: The distributions of the completed tasks (left) and hiring decisions (right) per employer in our dataset.

4.2 Feature engineering

The framework described in Section 3 assumes a feature vector $\boldsymbol{X}$, which captures the distinctive characteristics of the employer, the opening, and the contractor at hand. To construct this vector we draw on the rich previous literature on hiring decisions in online labor markets.

The first attributes that we consider capture the contractor’s reputation. Researchers have shown that higher feedback scores increase the hiring probability (e.g., [33]). In our setting, we observe an over-saturation of very reputable applicants; For example, the average feedback score of a contractor in our dataset is 4.77 out of 5.0, while the median is 4.92 out of 5.0. Even though inflated, feedback scores might still provide a noisy signal regarding the contractor’s quality. One of the possible explanations for reputation inflation is the phenomenon of ‘customer death’ [21]: Users who receive low feedback scores usually abandon the platform, and they either create new accounts and rejoin, or find alternative platforms to use. As a result, the total number of completed jobs as well as the total number of work-hours of a contractor on the platform are two additional signals of reputation. That is, a contractor who has been employed multiple times in the past is all else being equal more reputable than a contractor who has yet to be employed (similarly with work-hours). Hence, we start the construction of vector $\boldsymbol{X}$ by including these three signals: feedback score, number of completed jobs, and the total number of work-hours.

Next, we know from previous works that the contractors’ bidding prices are correlated with the outcomes of hiring decisions [12, 33, 37]. Specifically we know that high-value projects create greater rents for low-quality bidders, encouraging them to participate even when their likelihood of success is small [37]. In our study, we deal with two types of tasks, i.e., hourly and fixed priced. To capture the bidding effect along with its task-specific incremental effect we include in the feature vector (1) the bidding price, (2) a dummy variable that captures whether or not we have a fixed price task, and (3) an interaction term between the bid and the dummy fixed price variable.

Employers show a strong preference for contractors from their own country [11, 14]. The underlying mechanism that explains this is that communication is easier and trust increases when interacting

	Mean	Median	StD	Min	Max
Feedback score	4.77	4.92	0.39	1	5
Completed jobs	4.94	0	14.45	0	403
Total hours worked	563	40	1449	0	37697
Bid ($)	72.1	12.22	254	3	5000
Fixed price task	0.53	1	0.5	0	1
Bid x Fixed price	126	35	340	3	5000
Countries PMI	-0.49	-0.46	0.42	-3.68	4.44
Certifications PMI	2.34	2.08	1.73	-0.71	7.33
Skills inner product	1	1	1.2	0	18
Contractor invited	0.16	0	0.36	0	1
Order of application	26	15	33	0	291

Table 1: Descriptive statistics of the eleven attributes in feature vector X.

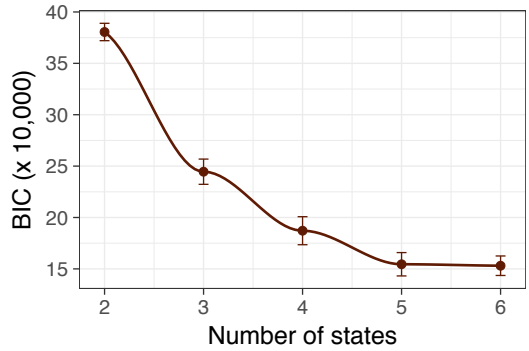

Figure 5: The BIC scores for different number of states (K) for our HMM. The error bars represent 95% confidence intervals. Based on the results, we pick $K = 5$.

with local contractors [4, 29]. To better capture these employer-contractor preferences, we use a location metric that calculates the point-wise mutual information (PMI) of the employer's and the contractor's countries [27]. PMI is a measure of how much the probability of a particular co-occurrence between two events differs from what we would expect to see if the two events were independent. PMI takes both positive and negative values; it is zero when the two events are independent, and it maximizes when the two events are perfectly associated. Formally [10]:

$$\text{Countries PMI}(C_a, C_i) = \log \frac{\Pr(C_a, C_i)}{\Pr(C_a)\Pr(C_i)} \ , \tag{18}$$

where C_a is the country of the hired contractor and C_i the country of the employer.

Online labor markets offer a series of skill tests and certifications. Goes and Lin [16] showed that certifications might have a negative impact on the likelihood of getting hired. In our work, we argue that not all certifications are the same; some certifications might be very important for certain jobs but completely irrelevant to others. We propose to capture the importance of a given certificate through its co-occurrence with certain types of jobs. Similarly to before, we use the point-wise mutual information (PMI) between the set of required skills by the job opening (O_o) and the set of certifications of the contractor (Q_a):

$$\text{Certifications PMI}(O_o, Q_a) = \log \frac{\Pr(O_o, Q_a)}{\Pr(Q_a)\Pr(O_o)} \ . \tag{19}$$

Furthermore, we estimate the inner product between the set of self-reported skills of the contractor and the set of required skills by the opening.

We complete the feature vector $\boldsymbol{X}$ by including the order of application (i.e., a measure of when did the candidate apply for the job), as well as whether the employer actively sought out and contacted the candidate to apply. We expect these two attributes to have predictive information, since the first one is a noisy signal of competition for the task, while the second one shows that the employer is preferential towards the invitee.

The complete feature vector $\boldsymbol{X}$ consists of eleven attributes. Table 1 shows the descriptive statistics of these features. Note that some variables have missing values: Not every contractor has previously worked on the platform, and as a result these contractors do not have an assigned feedback score. Similarly, some combinations of the certifications PMI and the countries PMI are not found in our training sets.[6] We impute these missing values to the median within task-category scores. In addition, some of the variables have really long tails (e.g., "Total hours worked"), and as a result we log-transform them. Finally, to accelerate the convergence of the optimization algorithm, we standardize all the non-binary attributes.

4.3 Evaluation setup and parameter estimation

For our analysis, we split our dataset into training (66%) and test sets (34%) based on employers (i.e., the sequence of outcomes of each employer can only be in either the test or the training set, but not in both). We use our training set to estimate the parameters of the HMM, as well as the parameters of the baseline models we consider (discussed later). We then use the test set to evaluate the performance of our approach. In the next paragraphs we describe the process of building our HMM and our baselines, and we conclude by discussing the evaluation metrics that we use.

HMM model selection: We first use the training set to empirically identify the number of unobserved states K that best explain our set of observations. Specifically, we estimate the Bayesian information criterion (BIC) scores for 1,000 random initializations, for each $K \in 2, 3, \ldots, 6$. We show the results in Figure 5. We see that $K = 5$ and $K = 6$ yield the lowest BIC scores. Since $K = 6$ is not statistically significantly better than $K = 5$, we choose the latter. Note that this selection process is common practice in similar scenarios [7, 28, 34].

Given that the number of states for the HMM ($K = 5$), we estimate the set of parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ that maximize the likelihood of Equation 17. Our optimization process depends on the initialization of these parameters, and as a result it is prone to stuck in local maxima. To increase the likelihood of reaching a potential global

[6] To avoid overfitting, we estimate the PMI scores for the exams and the countries on a completely separate dataset of 20,000 randomly chosen hiring decisions, that we do not use anywhere else in this study. Since some combinations in our focal dataset are not present in the PMI dataset, we end up with the missing values shown in Table 1.

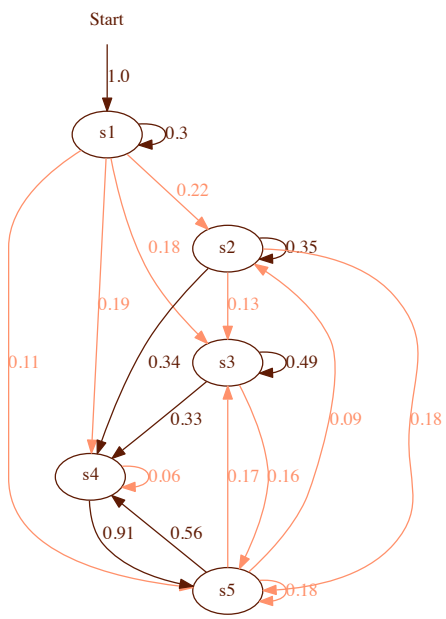

Figure 6: Our best-fitted HMM according to BIC score for $K = 5$. The transition probabilities are estimated through simulating employer decisions in the training set. We omit the emission probabilities to increase the readability of the graph. For the same reason, we only show transitions with probabilities greater than 0.05.

maximum, we conduct a search of 10,000 randomly selected prior configurations. We then choose the configuration that yields the lowest BIC score.

In Figure 6 we show the best-fitted HMM (according to BIC) on our training set. To estimate the transition probabilities, we simulate the hiring behaviors of employers in our training set (i.e., we simulate the paths of employers, and then average the resulting transition probabilities from each state). In the Figure, we distinguish the two layers shown in Figure 2: From state s_1 employers move to states in the second layer, $s_2 - s_5$. Once there, employers transition to states within the bottom layer, according to the outcomes of their hiring decisions and the estimated parameters $\boldsymbol{\gamma}$.

Baseline mmodels: To compare the performance of our HMM we use the previously described models for recommending candidates in online labor markets by Kokkodis *et al.* [27]. Specifically, we build a logistic regression, and a Bayesian network. Compered to their previous implementations in [27], we now formulate them to predict the outcomes in $\mathcal{Y}$. For the multi-class logistic regression, we estimate the following:

$$\Pr(Y_t|\boldsymbol{X}_t) = \text{softmax}(\boldsymbol{\delta}\boldsymbol{X}_t) \ . \tag{20}$$

Similarly, for the Bayesian network approach we get:

$$\Pr(Y_t|\boldsymbol{X}_t) = \frac{\prod_{x \in \{\boldsymbol{X}_t \cup Y_t\}} \Pr(x|pa(x))}{\Pr(Y_t)} \ , \tag{21}$$

where $pa(x)$ is the parent node of x in our network. As with the HMM framework, we use our training set to estimate the parameters of these models. For the Bayesian network, we identify its structure through the BIC approximation ([34]).

Evaluation metrics: In our scenario we are interested in rankings of applicants according to their likelihood of getting hired and doing a good job. As a result, the concept of accuracy as the fraction of corrected classified instances is both irrelevant and misleading: a majority classifier that always predicts "no hire" would have an accuracy of 93.6%.

An appropriate evaluation metric for our problem that (1) overcomes the skewed input distribution bias but also (2) evaluates the probabilistic ranking of the outcomes is the Area Under the ROC Curve[7] (AUC). Intuitively, the AUC score is the probability that our model will rank applicants correctly, according to their likelihood of (1) not being hired ("no hire"), (2) being hired and doing a bad job ("Negative outcome"), and (3) being hired and doing a good job ("Positive outcome").

To directly compare the performance of our HMM with the performance of the two baseline models we further estimate the AUC improvement percentage:

$$\text{AUC Improvement (\%)} = \frac{\text{AUC}_{HMM} - \text{AUC}_{baseline}}{\text{AUC}_{baseline}} * 100 \ . \tag{22}$$

4.4 Results:

In Figure 7 we show the AUC scores for our three models, across the three outcomes. A random predictor has an AUC=0.5 [36]. Our HMM framework works better than the other two algorithms in two out of the three classes we consider. Specifically, it yields an AUC of 0.76 in ranking applicants according to their likelihood of emitting a positive outcome, compared 0.68 and 0.66 of the logistic regression and the Bayesian network respectively. Similarly, our HMM outperforms the other two in terms of ranking instances according to their likelihood of not getting hired (0.76 vs. 0.67 and 0.65). To the contrary, our framework fails to accurately rank negative outcomes, where logistic regression seems to provide a descent ranking, with an AUC score of 0.65.

Up to this point, we showed that our framework outperforms the two alternative algorithms in two out of our three classes. Recall that the HMM captures the evolution of employers through consequent hiring decisions. Hence we should expect it to perform better, as employers make repeated hiring decisions. To compare the performance of the three models on consecutive hires, we estimate the AUC percentage improvement (Equation 22) for n or more completed tasks, where $n = \{0, 1, ..., 10\}$. When $n = 0$, we evaluate the performance of these approaches on the complete test set (i.e.,, same evaluation as the one presented in Figure 7). When $n = 1$, we evaluate only on the hiring decisions that followed the completion of the employer's first task. When $n = 2$, we evaluate only on the instances in our test set that followed the completion of the employer's second task, etc. We show the results in Figure 8. For the negative outcome, our HMM starts badly, presenting a negative improvement in the complete dataset (around 20% worse than the logistic regression and the Bayesian network). However,

[7]For a detailed analysis on the use of AUC see [36].

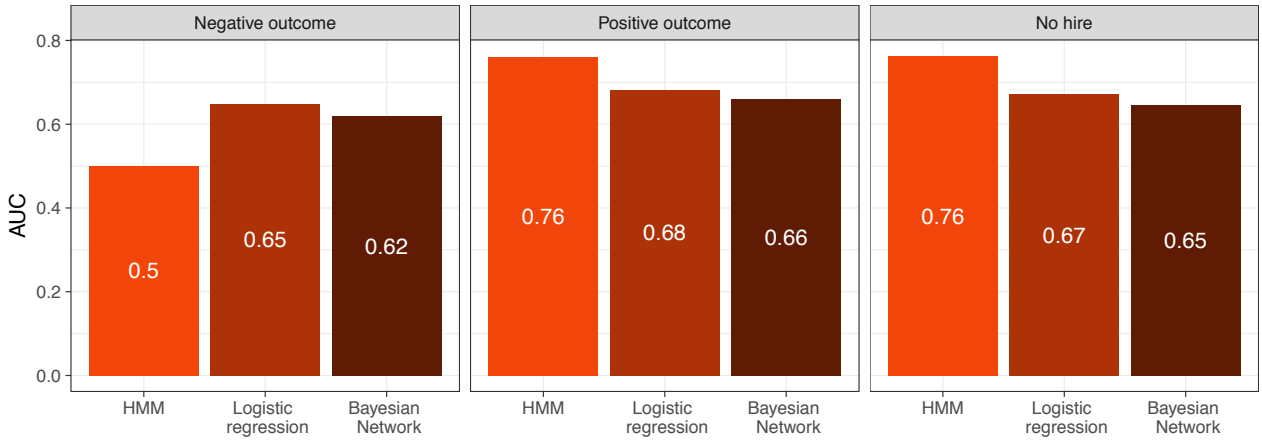

Figure 7: The AUC scores for the HMM framework and the two alternative static modeling approaches, Logistic regression and Bayesian network. We see that in "positive outcomes" and in "no hires" the HMM performs significantly better than the other two. To the contrary, the HMM fails to accurately rank "negative outcomes."

Figure 8: AUC improvement (Equation 22) of our approach over the two alternatives models, logistic regression and Bayesian network. Overall, we see that as the number of completed tasks increases, the improvement of our HMM increases. Only exception is the case of "positive outcomes", where the improvement ranges between 1% and 13%.

after the first set of completed tasks, our HMM recovers, and we see a positive improvement for $n \geq 4$. In fact, this improvement reaches 24% for $n = 10$. For positive outcomes, we do not observed a clear increasing trend of the improvement. However, the improvement over both alternative algorithms remains positive and ranges between 1% and 13%. Finally, for ranking applicants who are not getting hired, our HMM approach shows a slightly increasing trend, and it reaches an improvement of 22% for $n = 6$ and $n = 10$.

5 DISCUSSION

In this work we built and evaluated a structured HMM framework that captures the dynamic hiring behavior of employers in online labor markets. We showed that our framework has a strong predictive performance, and that it overall outperforms static implementations for ranking applicants.

The deployment of approaches that recommend candidates is critical for online labor markets. Such recommender frameworks (1) accelerate the employers' selection process and reduce search cost [5], and (2) they can discourage contractors from flooding openings by applying to jobs that have very low hiring probability. The observed improvement of our approach over the two alternative algorithms can have a strong impact on the marketplace: Improved matchings can increase both the employers' and the contractors' satisfaction, and as a result increase the likelihood of these users to keep using the platform.

The deployment of approaches that recommend candidates is critical for online labor markets. Such recommender frameworks

(1) accelerate the employers' selection process and reduce search cost [5], and (2) they can discourage contractors from flooding openings by applying to jobs that have very low hiring probability. The observed improvement of our approach over the two alternative algorithms can have a strong impact on the marketplace: Improved matchings can increase both the employers' and the contractors' satisfaction, and as a result increase the likelihood of these users to keep using the platform.

Methodological contributions: Besides the benefits of the implementation of our approach, our study provides a technically solid framework for problems that are broadly similar. For example, a version of our approach could be used in recommending applicants in the offline setting: By analyzing career trajectories of candidates on LinkedIn we could build a similar dynamic recommendation framework. For such an approach, the observed outcomes could be "upward trajectory" and "downward trajectory" (since we do not observe "no hire" outcomes on LinkedIn).

Even further, our framework can be used in vastly different scenarios of sequential observations. For instance, we could use our approach to estimate the expertise of a worker on a given set of skills. The idea is that contractors evolve and learn new skills all the time. In addition, because feedback scores are inflated, it is very hard–if not impossible–for employers to assess the true expertise of a worker on a given set of skills. By using signals of the contractors' histories on the platforms (e.g., what types of tasks they work on, what are the task outcomes, etc.) we can dynamically provide estimates for the latent expertise of a contractor on a given set of skills.

Conclusions: To conclude, our work is the first to propose a framework that explicitly models the dynamic and heterogeneous nature of employers. Through our evaluation on a dataset of online hiring decisions we show that our approach outperforms the static implementations of recommending candidates. Our proposed framework has the potential to minimize the employers' search cost and increase the market's transaction volume through the overall increase in satisfaction of the contractors and the employers.

REFERENCES

[1] Ajay Agrawal, John Horton, Nicola Lacetera, and Elizabeth Lyons. 2015. Digitization and the contract labor market. *Economic Analysis of the Digital Economy* (2015), 219.

[2] Ajay K Agrawal, Nicola Lacetera, and Elizabeth Lyons. 2013. *Does information help or hinder job applicants from less developed countries in online markets?* Technical Report. National Bureau of Economic Research.

[3] George A Akerlof. 1970. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* (1970), 488–500.

[4] Ashish Arora and Chris Forman. 2007. Proximity and information technology outsourcing: How local are IT services markets? *Journal of Management Information Systems* 24, 2 (2007), 73–102.

[5] J Yannis Bakos. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science* 43, 12 (1997), 1676–1692.

[6] Rajiv D Banker and Iny Hwang. 2008. Importance of measures of past performance: Empirical evidence on quality of e-service providers. *Contemporary Accounting Research* 25, 2 (2008), 307–337.

[7] C.M. Bishop et al. 2006. *Pattern recognition and machine learning*. Vol. 4. springer New York.

[8] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16, 5 (1995), 1190–1208.

[9] Jason Chan and Jing Wang. 2017. Hiring preferences in online labor markets: Evidence of a female hiring bias. *Management Science (forthcoming)* (2017).

[10] Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16, 1 (1990), 22–29.

[11] David Gefen and Erran Carmel. 2008. Is the world really flat? A look at offshoring at an online programming marketplace. *MIS Quarterly* (2008), 367–384.

[12] David Gefen, Gavriel Gefen, and Erran Carmel. 2015. How project description length and expected duration affect bidding and project success in crowdsourcing software development. *Journal of Systems and Software* (2015).

[13] Tomer Geva and Maytal Saar-Tsechansky. 2016. Who's a good decision maker? Data-driven expert worker ranking under unobservable quality. *International Conference on Information Systems (ICIS)* (2016).

[14] Ejaz Ghani, William R Kerr, and Christopher Stanton. 2014. Diasporas and outsourcing: evidence from oDesk and India. *Management Science* 60, 7 (2014), 1677–1697.

[15] Anindya Ghose, Param Vir Singh, and Vilma Todri. 2017. Got Annoyed? Examining the Advertising Effectiveness and Annoyance Dynamics. In *International Conference on Information Systems (ICIS)*.

[16] Paulo Goes and Mingfeng Lin. 2012. Does information really unravel? Understanding factors that motivate sellers to seek third-party certifications in an online labor market. *NET Institute Working Paper* 12-02 (2012).

[17] Yili Hong and Paul A Pavlou. 2017. On buyer selection of service providers in online outsourcing platforms for IT services. *Information Systems Research* (2017).

[18] John Horton and Joseph Golden. 2015. Reputation inflation: evidence from an online labor market. *Working Paper, NYU* (2015).

[19] John Joseph Horton and Ramesh Johari. 2015. At what quality and what price?: Eliciting buyer preferences as a market design problem. *Proceedings of the Sixteenth ACM Conference on Economics and Computation* (2015), 507–507.

[20] N. Hu, J. Zhang, and P.A. Pavlou. 2009. Overcoming the J-shaped distribution of product reviews. *Commun. ACM* 52, 10 (2009), 144–147.

[21] Kinshuk Jerath, Peter S Fader, and Bruce GS Hardie. 2011. New perspectives on customer "death" using a generalization of the Pareto/NBD model. *Marketing Science* 30, 5 (2011), 866–880.

[22] Marios Kokkodis. 2012. Learning from positive and unlabeled Amazon reviews: Towards identifying trustworthy reviewers. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*. ACM, 545–546.

[23] Marios Kokkodis and Panos Ipeirotis. 2014. The utility of skills in online labor markets. *Proceedings of the AIS International Conference on Information Systems (ICIS)* (2014).

[24] Marios Kokkodis and Panagiotis G Ipeirotis. 2013. Have you done anything like that?: predicting performance using inter-category reputation. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining (WSDM)*. ACM, 435–444.

[25] Marios Kokkodis and Panagiotis G Ipeirotis. 2015. Reputation transferability in online labor markets. *Management Science* 62, 6 (2015), 1687–1706.

[26] Marios Kokkodis and Theodoros Lappas. 2016. Realizing the Activation Potential of Online Communities. In *International Conference on Information Systems (ICIS)*.

[27] Marios Kokkodis, Panagiotis Papadimitriou, and Panagiotis G Ipeirotis. 2015. Hiring behavior models for online labor markets. In *Proceedings of the eighth ACM International Conference on Web Search and Data Mining (WSDM)*. ACM.

[28] Daphne Koller and Nir Friedman. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

[29] Gueorgi Kossinets and Duncan J Watts. 2009. Origins of homophily in an evolving social network. *Amer. J. Sociology* 115, 2 (2009), 405–450.

[30] Mingfeng Lin, Yong Liu, and Siva Viswanathan. 2018. Effectiveness of Reputation in Contracting for Customized Production: Evidence from Online Labor Markets. *Management Science* 64, 1 (2018), 345–359.

[31] Mingfeng Lin and Siva Viswanathan. 2015. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* 62, 5 (2015), 1393–1414.

[32] Roy Mill. 2011. Hiring and learning in online global labor markets. *Available at SSRN 1957962* (2011).

[33] Antonio Moreno and Christian Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research* 25, 4 (2014), 865–886.

[34] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. The MIT Press.

[35] Amanda Pallais. 2014. Inefficient hiring in entry-level labor markets. *The American Economic Review* 104, 11 (2014), 3565–3599.

[36] F. Provost and T. Fawcett. 2001. Robust classification for imprecise environments. *Machine Learning* 42, 3 (2001), 203–231.

[37] Eli M Snir and Lorin M Hitt. 2003. Costly bidding in online markets for IT services. *Management Science* 49, 11 (2003), 1504–1520.

[38] Christopher Stanton and Catherine Thomas. 2015. Landing the first job: The value of intermediaries in online hiring. *The Review of Economic Studies* (2015), rdv042.

[39] Hema Yoganarasimhan. 2013. The value of reputation in an online freelance marketplace. *Marketing Science* 32, 6 (2013), 860–891.