

## A Joint Learning Approach to Intelligent Job Interview Assessment

Dazhong Shen<sup>1,2</sup>, Hengshu Zhu<sup>2,\*</sup>, Chen Zhu<sup>2</sup>, Tong Xu<sup>1,2</sup>, Chao Ma<sup>2</sup>, Hui Xiong<sup>1,2,3,4,\*</sup>

<sup>1</sup>Anhui Province Key Lab of Big Data Analysis and Application, University of S&T of China,

<sup>2</sup>Baidu Talent Intelligence Center,

<sup>3</sup>Business Intelligence Lab, Baidu Research,

<sup>4</sup>National Engineering Laboratory of Deep Learning Technology and Application, China.

sdz@mail.ustc.edu.cn, {zhuhengshu, zhuchen02, machao13}@baidu.com,

tongxu@ustc.edu.cn, xionghui@gmail.com

### Abstract

The job interview is considered as one of the most essential tasks in talent recruitment, which forms a bridge between candidates and employers in fitting the right person for the right job. While substantial efforts have been made on improving the job interview process, it is inevitable to have biased or inconsistent interview assessment due to the subjective nature of the traditional interview process. To this end, in this paper, we propose a novel approach to intelligent job interview assessment by learning the large-scale real-world interview data. Specifically, we develop a latent variable model named **Joint Learning Model on Interview Assessment (JLMIA)** to jointly model job description, candidate resume and interview assessment. JLMIA can effectively learn the representative perspectives of different job interview processes from the successful job interview records in history. Therefore, a variety of applications in job interviews can be enabled, such as person-job fit and interview question recommendation. Extensive experiments conducted on real-world data clearly validate the effectiveness of JLMIA, which can lead to substantially less bias in job interviews and provide a valuable understanding of job interview assessment.

### 1 Introduction

As one of the most important functions in human resource management, talent recruitment aims on acquiring the right talents for organizations and always has direct impact on business success. As indicated in an article from Forbes, US corporations spend nearly 72 billion dollars each year on a variety of recruiting services, and the worldwide amount is likely three times bigger [Bersin, 2013]. In particular, job interview, which is considered as one of the most useful tools and the final testing ground for evaluating potential employees in the hiring process, has attracted more and more attentions in human resource management. While substantial efforts have been made on the improvement of job interview

process, traditional interview process has a substantial risk of bias due to the subjective nature of the process. This situation could be even more severe, since different interviewers may have different technical backgrounds or different experience levels in personal qualities. This may lead to a biased or incomplete assessment of job candidate.

Recently, the Artificial Intelligence (AI) trend has made its way to talent recruitment, such as job recommendation [Malinowski *et al.*, 2006; Paparrizos *et al.*, 2011; Zhang *et al.*, 2014], talent mapping [Xu *et al.*, 2016], and market trend analysis [Zhu *et al.*, 2016]. However, fewer efforts have been made on enhancing the quality and experience of job interview. A critical challenge along this line is how to reveal the latent relationships between job position and candidate, and further form perspectives for effective interview assessment. Intuitively, experienced interviewers could discover the topic-level correlation between job description and resume, and then design the interview details to measure the suitability of applicants. For example, a candidate for “Software Engineer”, who has strong academic background, might be interviewed with questions not only about “Algorithm”, “Programming”, but also “Research”. Meanwhile, compared with the technical interview, the vocabulary of comprehensive interview could be largely different.

To this end, we propose a novel approach to intelligent job interview assessment by learning the large-scale real-world interview data. Specifically, we develop a latent variable model named **Joint Learning Model on Interview Assessment (JLMIA)** to jointly model job description, candidate resume and interview assessment. JLMIA can effectively learn the representative perspectives of different job interview processes from the historical successful job interview records. Also, two categories of interviews, *technical* and *comprehensive* interviews, which are hosted by technical and managerial interviewers respectively, could be well differentiated. Furthermore, based on JLMIA, we also provide solutions for two applications named person-job fit and interview question recommendation. Extensive experiments conducted on real-world data clearly validate the effectiveness of JLMIA, which can lead to substantially less bias in job interviews and provide a valuable understanding of job interview assessment.

\*Corresponding Author.

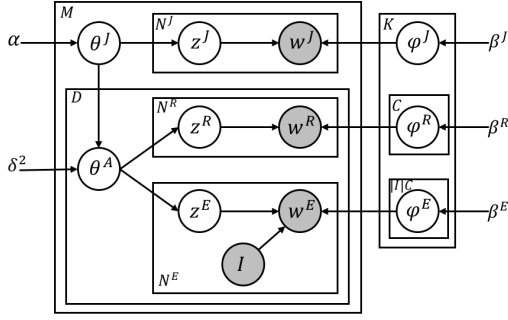


Figure 1: The graphical representation of JLMIA.

## 2 Problem Statement

Formally, our data set contains the recruitment documents of  $|M|$  unique jobs, i.e.,  $S = \{S_m = (J_m, A_m)\}_{m=1}^{|M|}$ , where  $J_m$  is the job description of the  $m$ -th job and  $A_m$  is the interview records of this job. Specifically,  $A_m = \{(R_{md}, E_{md})\}_{d=1}^{|D_m|}$  contains  $|D_m|$  interviews, where  $R_{md}$  is the resume of candidate in  $d$ -th interview, and  $E_{md}$  is the corresponding interview assessment. Since all of the job descriptions, resumes, and interview assessments are textual data, we use bag-of-words to represent them, e.g.,  $J_m = \{w_{mj}^J\}_{j=1}^{N_m^J}$ , similar to  $R_{md}$  and  $E_{md}$ .

A job description  $J_m$  contains detailed job requirements, and a resume  $R_{md}$  mainly consists of the past experiences of this candidate that can reflect her abilities. Meanwhile, the evaluation about a candidate in interview assessments bridges the gap between job requirements and her ability. And according to the goal of interviews, interview assessments can be further divided into technical and comprehensive interview.

As we known, during the interview, interviewers tend to ask questions related to the work experiences of candidates. Thus there often exists strong correlation between interview assessments and resumes. However, job description is usually more abstract than resumes, and candidates with different backgrounds may be suitable for the same job. Thus we think although there exists correlation between job descriptions and resumes, the diversity of job descriptions is less than that of resumes. In addition, it is obvious that the focus of interviews is different according to their goals. Thus it is better to model the differences between technical and comprehensive interview.

Generally, the main tasks in this paper can be summarized as: **Task 1**, how to discover the strong correlation between resumes and interview assessments? **Task 2**, how to model the latent relationships between job descriptions and resumes? **Task 3**, how to distinguish the differences between different interview categories?

## 3 Technical Details of JLMIA

To solve the above tasks, we propose a novel joint learning model, namely JLMIA. In this section, we will formally introduce its technical details.

### 3.1 Model Formulation

To model the latent semantics in job description, resume, and interview assessment, we assume there exist latent topics, represented by  $\varphi^J$ ,  $\varphi^R$  and  $\varphi^E$ , in all of them. And

### Algorithm 1: The Generative Process of JLMIA for Resume and Interview Assessment

1. For each topic  $k$  of candidate interview record:
  - (a) Draw  $\varphi_k^R$  from the Dirichlet prior  $Dir(\beta^R)$ .
  - (b) Draw  $\varphi_k^{ET}$  and  $\varphi_k^{EC}$  from the Dirichlet prior  $Dir(\beta^E)$ .
2. For each job description  $J_m$ :
  - (a) Sample topic distribution  $\theta_m^J \sim Dir(\alpha)$ .
3. For each candidate interview record pair  $(R_{md}, E_{md}, I_{md})$ :
  - (a) Sample topic distribution  $\theta_{md}^A \sim N(h(\theta_m^J, C), \delta^2 I)$
  - (b) For the  $r$ -th word  $w_{mdr}^R$  in resume  $R_{md}$ :
    - i. Draw topic assignment  $z_{mdr}^R \sim Multi(\pi(\theta_{md}^A))$ .
    - ii. Draw word  $w_{mdr}^R \sim Multi(\varphi_{z_{mdr}^R}^R)$ .
  - (c) For the  $e$ -th word  $w_{mde}^E$  in interview assessment  $E_{md}$ :
    - i. Draw topic assignment  $z_{mde}^E \sim Multi(\pi(\theta_{md}^A))$ .
    - ii. Draw word  $w_{mde}^E \sim Multi(\varphi_{z_{mde}^E}^{ET})$  ( $I_{mde} = TI$ ).
    - iii. Draw word  $w_{mde}^E \sim Multi(\varphi_{z_{mde}^E}^{EC})$  ( $I_{mde} = CI$ ).

our tasks are further transformed to model the relationships among these latent topics. First, to model the strong correlation between resume  $R_{md}$  and interview assessment  $E_{md}$ , we directly assume they share the same tuple-specific distribution  $\theta_{md}^A$  over topics. Second, for revealing the relationships between job descriptions and resumes along with the differences between their diversity, we generate  $\theta_{md}^A$  from the logistic-normal distribution with mean parameter related to the topic distribution of job description  $\theta_m^J$ . And the topic numbers of  $\varphi^J$ ,  $\varphi^R$  and  $\varphi^E$  are set as  $|k^J| = |k^R| = C \cdot |k^J| = CK$ . In other words, for each topic in  $\varphi^J$ , there are  $C$  topics in  $\varphi^R$  ( $\varphi^E$ ) related to it. Third, we use a label  $I \in \{TI, CI\}$  (e.g., **T**echnical **I**nterview or **C**omprehensive **I**nterview) to indicate the type of interview for each interview assessment, where different types of interview assessment are generated from different topics  $\varphi^E \in \{\varphi^{ET}, \varphi^{EC}\}$ . To simplify our model, we follow the idea in [Wang and McCallum, 2006], and set the interview label for each word in interview assessment instead of the entire interview assessment.

The graphical model of JLMIA is shown in Figure 1. Since the generative process of job description is the same as Latent Dirichlet Allocation (LDA) [Blei *et al.*, 2003], here we only list the generative process for resume and interview assessment  $A = \{A_m\}_{m=1}^{|M|}$ , showed in Algorithm 1, where  $h(\theta, C)$ , in line 3.(a), is a vector concatenating  $C$  log vectors of  $\theta$ , i.e.,  $h(\theta_m^J, C)_k = \log \theta_{m,k}^J$ ,  $k' = k \bmod K$ ,  $1 \leq k' \leq K$ , and  $\pi(\theta)$ , in line 3.(b).i and 3.(c).i, is the logistic transformation, i.e.,  $\pi(\theta_{md}^A)_k = \frac{\exp\{\theta_{md,k}^A\}}{\sum_{i=1}^{CK} \exp\{\theta_{md,i}^A\}}$ .

Due to the non-conjugacy of the logistic normal and multinomial, the latent parameters posterior is intractable. Thus we propose a variational inference algorithm for JLMIA.

### 3.2 Variational Inference for JLMIA

Here, we develop a variational inference algorithm for JLMIA based on mean-field variational families. The basic idea behind variational inference is to optimize the free para-

meters of a distribution over the latent variables, so that the distribution is close in Kullback-Liebler (KL) divergence to true posterior, which can be substituted. In our model, let us denote all latent variable parameters by  $\Phi$  and all hyper-parameters by  $\Omega$ . Following the generative process, the joint distribution can be factored as:

$$p(S, \Phi | \Omega) = p(\Phi | \Omega) \prod_{m=1}^{|M|} P(S_m | \Phi), \quad (1)$$

where each component can be calculated by:

$$\begin{aligned}
 p(S_m | \Phi) &= p(J_m | z_m^J, \varphi^J) \prod_{d=1}^{|D_m|} p(R_{md} | z_{md}^R, \varphi^R) p(E_{md} | z_{md}^E, \varphi^E, I_{md}), \\
 p(\Phi | \Omega) &= \prod_{m=1}^{|M|} \prod_{d=1}^{|D_m|} p(\theta_{md}^A | \theta_m^J, \delta^2) \prod_{r=1}^{N_{md}^R} p(z_{mdr}^R | \theta_{md}^A) \prod_{e=1}^{N_{md}^E} p(z_{mde}^E | \theta_{md}^A) \\
 &\quad \times \prod_{k=1}^K p(\varphi_k^J | \beta^J) \prod_{k=1}^{CK} p(\varphi_k^R | \beta^R) p(\varphi_k^{ET}, \varphi_k^{EC} | \beta^E) \\
 &\quad \times \prod_{m=1}^{|M|} p(\theta_m^J | \alpha) \prod_{j=1}^{N_m^J} p(z_{mj}^J | \theta_m^J).
 \end{aligned}$$

Then, corresponding to this joint distribution, we posit the fully factorized variational families as following, where the detail description of each term can be found in Appendix:

$$\begin{aligned}
 q(\Phi) &= \prod_{k=1}^K q(\varphi_k^J) \prod_{k=1}^{CK} q(\varphi_k^R) q(\varphi_k^{ET}) q(\varphi_k^{EC}) \prod_{m=1}^M q(\theta_m^J) \prod_{j=1}^{N_m^J} q(z_{mj}^J) \\
 &\quad \times \prod_{m=1}^M \prod_{d=1}^{|D_m|} \prod_{k=1}^{CK} q(\theta_{md,k}^A) \prod_{r=1}^{N_{md}^R} q(z_{mdr}^R) \prod_{e=1}^{N_{md}^E} q(z_{mde}^E). \quad (2)
 \end{aligned}$$

According to [Blei *et al.*, 2017], minimizing the KL divergence between variational distribution and true posterior, is equivalent to maximize the log likelihood bound of job interview records, which is the evidence lower bound (*ELBO*):

$$\begin{aligned}
 \log p(S | \Omega) &\geq E_q[\log p(S, \Phi | \Omega)] + H(q) \\
 &= E_q[\log p(\Phi | \Omega)] + \sum_{m=1}^{|M|} E_q[\log p(S_m | \Phi)] + H(q), \quad (3)
 \end{aligned}$$

where the expectation  $E_q[\cdot]$  is taken with respect to the variational distribution in Equation 2, and  $H(q)$  denotes the entropy of that distribution.

The largest challenge to maximize *ELBO* is the non-conjugacy of logistic normal and multinomial, which leads to the difficulty in computing the expected log probability of topic assignments in documents of each candidate interview records. Similar to [Wang and Blei, 2011], we introduce a new variational parameter  $\zeta = \{\zeta_{m1:|D_m|}\}_{m=1:|M|}$  to preserve the lower bound of *ELBO*. Here we take the  $E_q[\log p(z_{mdr}^R | \theta_{md}^A)]$  as an example to explain it (the  $E_q[\log p(z_{mde}^E | \theta_{md}^A)]$  can be computed in a similar way):

$$\begin{aligned}
 E_q[\log p(z_{mdr}^R | \theta_{md}^A)] &= E_q[\theta_{md, z_{mdr}^R}^A] - E_q[\log(\sum_{k=1}^{CK} \exp\{\theta_{md,k}^A\})] \\
 &\geq E_q[\theta_{md, z_{mdr}^R}^A] - \zeta_{md}^{-1} (\sum_{k=1}^{CK} E_q[\exp\{\theta_{md,k}^A\}] + 1 - \log(\zeta_{md})). \quad (4)
 \end{aligned}$$

For maximizing the *ELBO*, we develop an EM-style algorithm with coordinate ascent approach to optimize parameters, the details of which can be found in Appendix.

## 4 Application

Here, we will introduce two applications enabled by JLMIA, i.e., Person-Job Fit and Interview Question Recommendation.

### 4.1 Person-Job Fit

Person-Job Fit is the process of matching the right talent for the right job. Formally, given a job description  $J_g$  and a resume  $R_g$ , the objective is to measure their matching degree. Specifically, we need to first leverage JLMIA to infer the latent topic distributions of  $J_g$  and  $R_g$  respectively. However, our model cannot infer the topic distribution for an individual resume or job description. Thus we construct a  $S$  for resumes (job descriptions), where all of other data are set as empty, and infer the corresponding topic distribution.

After the variational parameters of topic assignment of each word,  $\phi_g^J$  and  $\phi_g^R$ , are learned, we can compute the document-topic distribution by:

$$\begin{aligned}
 \theta_{g,k}^J &= \frac{1}{N^J} \sum_{n=1}^{N^J} \phi_{gn,k}^J \quad k = 1, \dots, K, \\
 \theta_{g,k}^R &= \frac{1}{N^R} \sum_{n=1}^{N^R} \phi_{gn,k}^R \quad k = 1, \dots, CK. \quad (5)
 \end{aligned}$$

Then, by computing the similarity between  $\theta_g^J$  and  $\theta_g^R$ , we can measure the suitability between job description and resume. Actually, any distance calculation formulation between two probability distributions can be used here for measuring the similarity, such as Cosine distance and Kullback-Leibler divergence. Note that, since the dimension of  $\theta_g^J$  and  $\theta_g^R$  may be different, here we have:

$$\tilde{\theta}_{g,k}^R = \sum_{c \in C_k} \theta_{g,c}^R \quad k = 1, \dots, K,$$

where  $C_k$  is a set of mapping index that satisfies  $\theta_{g,c}^R$  ( $c \in C_k$ ) is generated from  $\theta_{g,k}^R$ . In particular, the vectors  $\theta_g^J$  and  $\theta_g^R$  learned by JLMIA can be regarded as low-rank semantic representations of job description and resume. Thus, using these representations instead of original bag-of-words as features for training a classifier (e.g., Random Forest) is another solution for Person-Job Fit.

### 4.2 Interview Question Recommendation

During the interview, interviewers need to ask some questions to evaluate candidates. However, due to the limited expert resources, sometimes the interviewers may not have enough domain knowledge to prepare discriminative questions for systematically judging the competencies of candidates, especially from the view of Person-Job Fit. Thus, in this paper, we propose an effective algorithm for recommending interview questions based on JLMIA and interview questions accumulated in historical assessments.

To be specific, given a question database  $Q = \{q_i\}_{i=1}^N$ , the problem of interview question recommendation is defined as retrieving a set of questions  $X$  that are related to a given query  $\Upsilon$  (i.e., job requirement item or experience item of candidate). Similar to the process of computing the topic distributions  $\theta_g^R$ , we can compute the topic distribution  $\theta_i^Q$  of each question  $q_i \in Q$ , through regarding interview questions as a part of interview assessment. Let  $\theta_i^Q$  denote a latent representation of question  $q_i$  and the latent representation of the given query  $\Upsilon$  as  $\theta_g^\Upsilon \in \{\theta_g^J, \theta_g^R\}$ .

To recommend high quality questions to interviewers, on the one hand, the selected question set  $X \subset Q$ ,  $|X| = L$  should be relevant to the query  $\theta_g^\Upsilon$ , on the other hand, we hope

to avoid making the questions in  $X$  too similar to each other. To balance relevance and diversity of selected question set, we select questions  $X$  by maximizing the following objective function:

$$F(\Upsilon, X) = \mu \frac{Rel(\Upsilon, X)}{\overline{Rel}} + (1 - \mu) \frac{Div(X)}{\overline{Div}}$$

$$= \mu \frac{\sum_{q_j \in X} Sim(\theta_j^\Upsilon, \theta_j^Q)}{\overline{Rel}} + (1 - \mu) \frac{\sum_{q_i \in X} \sum_{q_j \in X, q_j \neq q_i} Dis(\theta_i^Q, \theta_j^Q)}{\overline{Div}},$$

(6)

*s.t.*  $X \subset Q, |X| = L, 0 < \mu < 1,$

where,  $Rel(\Upsilon, X)$  and  $Div(X)$  measure the relevance and diversity above,  $Sim(*, *)$  is chosen as  $Cosine(*, *)$  while  $Dis(*, *)$  is set as  $1 - Cosine(*, *)$ , and  $\overline{Rel}$  and  $\overline{Div}$  are normalization factors, commonly chosen as the maximum possible values of  $Rel(\Upsilon, X)$  and  $Div(X)$  respectively.

In general, the calculation of addressing  $F(X)$  is computationally prohibitive, since we will suffer the assemble explosion problem if we calculate Equation 6 for all subsets  $X$ . Fortunately, since  $F(X)$  defined in this paper is submodule [Tang *et al.*, 2014], the simple greedy algorithm could achieve a  $(1 - 1/e)$  approximation of the optimal solution,

## 5 Experimental Results

In this section, we will introduce the performance of JLMIA based on a real-world interview data set.

### 5.1 Experimental Setup

The data set used in the experiments is the historical recruitment data provided by a high-tech company in China, which contains total 14,702 candidate interview records. To be specific, with the help of several staffing experts, we manually screened records with high quality interview assessment written by senior interviewers, and removed the records which lack details in job description or resume. After that, the filtered data set contains 4,816 candidate interview records related to 409 job positions. In JLMIA, we empirically set fixed parameters  $\{\delta^2, \beta^J, \beta^R, \beta^E\} = \{0.01, \mathbf{0.1}, \mathbf{0.1}, \mathbf{0.1}\}$ . Note that, our model is trained with original Chinese words. And for facilitating demonstration, all experimental results were translated into English.

### 5.2 Evaluation of Topic Joint Learning

To evaluate the effectiveness of joint learned topics by JLMIA, we first trained our model on all successful job interview records. In particular, we set the parameters  $K = 10$  and  $C = 2$ . Table 1 shows one randomly selected latent topic of job description and corresponding topics of resume and two types of interview assessments. Each topic is represented by several words with the highest probability.

We can observe that the topic of job description, containing “Experience” and “Foundation” of “PHP” and “Web”, should be related to web development. Similarly, the corresponding topics of resume and technical interview assessment also contain front-end-related keywords, “HTTP”, “Web Site”, “Element”, “JS” and “CSS”, which indicate the professional skills of candidates. Thus we believe that our model can effectively reveal the latent relationship among job description, resume and interview assessment. More interestingly, we can find that topic #1 and topic #11 of resume, which are both generated

Job Description	Resume	Tech. Interview	Com. Interview
Topic 1	Topic 1	Topic 1	Topic 1
Experience	Function	Foundation	Technology
Foundation	Management	Knowledge	Communication
Technology	Backstage	Code	Study
PHP	HTTP	Element	Knowledge
Web	Moudle	Development	Development
Proficient	Topic 11	Topic 11	Topic 11
Engineer	Web Site	JS	Job
Interest	Web Page	Methods	Pressure
Development	System	CSS	Solution
Web Page	Web	Elements	Like
Maintenance	Framework	Events	Work Overtime

Table 1: Topic example of JLMIA

from topic #1 of job description, contain different keywords, which validate the assumption that the diversity of job description is less than that of resume. Meanwhile, compared with technical interview assessment, there are more keywords like “Communication”, “Pressure” or “Work Overtime” in comprehensive interview assessment, which are related to the evaluation of personal qualities.

### 5.3 Performance of Person-Job Fit

Here, we evaluate the performance of JLMIA in terms of Person-Job Fit. Specifically, given a job description and a resume, we treat their latent topic distributions learned by our model as their representation vectors. Then, we train classic classifiers to predict the matching degree between the job and the candidate. Besides, to further demonstrate the effectiveness of our model, we also use the similarities between their representation vectors for measuring Person-Job Fit.

#### Benchmark Methods

We selected Latent Dirichlet Allocation (LDA) and bag-of-words (BOW) vector representation as baselines. For LDA, we merged the resume of a candidate and the job she applied for as a document for learning the latent topics. And for bag-of-words, where the  $i$ -th dimension of each vector is the frequency of the  $i$ -th word of the vocabulary, it itself is a kind of representation. Due to the limited space and similar trends of results, here we only selected Cosine and Kullback-Leibler similarity based approaches, and selected Random Forests and GBDT as classifiers. Please note that because the similarity between two BOW vectors is meaningless, we did not treat it as a baseline here.

#### Data Preparation

Different from the similarity based approaches, only one type of samples, i.e., positive samples, is required, the classifier based approaches need to prepare unsuitable pairs of job description and resume as negative samples to train classifiers. Although we can intuitively regard the historical failed job applications as negative samples, we do not know the exact reasons behind these failures. For example, some failed applications are just due to the low pay benefits, or other similar reasons in offer negotiation. Therefore, we manually generated the same number of negative samples to train classifier by randomly selecting resumes and job descriptions from the successful job interview records. Along this line, the experiments will only focus on the representation of latent topics, while interference from other factors will be impaired. After that, we randomly selected 80% data for model training and the other 20% data for test.

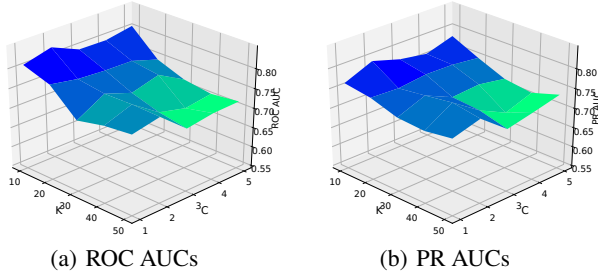


Figure 2: The Person-Job Fit performance of JLMIA based on Cosine similarity and different parameters.

		ROC AUC	PR AUC
Cosine Similarity	JLMIA	<b>0.8279</b>	<b>0.7935</b>
	LDA	0.7026	0.7223
Kullback-Leibler Divergence	JLMIA	<b>0.8234</b>	<b>0.8094</b>
	LDA	0.6589	0.6579
Random Forest (n_estimators=400)	JLMIA	<b>0.9012</b>	<b>0.8975</b>
	LDA	0.7359	0.7341
	BOW	0.6716	0.6761
GBDT (n_estimators=100, max_depth=9)	JLMIA	<b>0.8564</b>	<b>0.8311</b>
	LDA	0.7092	0.6810
	BOW	0.6531	0.6723

Table 2: The Person-Job Fit performance of different approaches.

### Performance Analysis

To evaluate the parameter sensitivity, we trained JLMIA by varying the parameter  $K$  from 10 to 50, and the parameter  $C$  from 1 to 5. The person-job fit performance of JLMIA based on Cosine similarity and different parameters is shown in Figure 2(a) and 2(b). We can find that the Receiver Operating Characteristic (ROC) AUCs and Precision-Recall (PR) AUCs are both better with small  $K$ , and reach the highest with  $K = 10$  and  $C = 2$ . Therefore, we chose the best parameters  $K$  and  $C$  for the following experiments. Similarly, we also evaluated LDA model with different topic number parameters  $K$ , and chose  $K = 30$  for other experiments.

Table 2 shows the Person-Job Fit performances of JLMIA and baselines. From the results, we find that our model consistently outperforms other baselines in both similarity based approaches and classifier based approaches. It indicates that JLMIA can effectively capture the latent relationship between job description and resume. More interestingly, the performances of JLMIA in similarity based approach is also higher than most of baselines. It clearly demonstrates the effectiveness of the representation learned by JLMIA.

### 5.4 Performance of Question Recommendation

To evaluate the performance of interview question recommendation of JLMIA, we first collected 1,085 interview questions as the candidate set from historical interview assessments, and then, compared JLMIA ( $K = 10$  and  $C = 2$ ) with BM25, a classic information retrieval model based on keywords matching which ignores the latent relationship between queries and questions. In our algorithm, the parameters are empirically set as  $Rel = 5$ ,  $Div = 20$  and  $\mu = 0.9$ .

We randomly selected 100 experience items as the queries. For each query, we recommend 10 questions by JLMIA and

	Relevance	Diversity	Personal Quality
JLMIA-TI	8.06	2.90	2.17
JLMIA-CI	7.72	2.84	3.22
BM25	7.14	1.67	1.00

Table 3: The question recommendation performance of JLMIA and BM25 with 10 questions recommended

Given experience item	I am familiar with <b>HTML</b> and <b>CSS programming</b> , and have some <b>web development</b> experience.
Questions recommended by JLMIA for technical interview	T1. What are <b>Ajax</b> and <b>Interactive Model</b> ? What are the differences between <b>Synchronous</b> and <b>Asynchronous requests</b> ? How to solve <b>Cross-domain issues</b> ?
	T2. What are the meanings of <b>Graceful Degradation</b> and <b>Progressive Enhancement</b> ?
	T3. How to make text centered vertically by <b>CSS programming</b> .
	T4. What is the role of the <b>HTTP status code</b> ?
Questions recommended by JLMIA for comprehensive interview	C1. Talk about <b>OSI</b> , <b>TCP/IP</b> and <b>Five-layers Network Model</b> .
	C2. What are the differences between <b>HTML</b> and <b>XHTML</b> ?
	C3. Do you think finding a job is not easy for you?
	C4. What are the differences between <b>Scrollbar</b> and <b>JScrollPane</b> ?
Questions recommended by BM25	B1. What are <b>web</b> applications?
	B2. Talk about your understanding of the semantics of <b>HTML</b> .
	B3. Please <b>program</b> a read-write lock with a normal mutex.
	B4. Talk about your understanding of the <b>web</b> standards and W3C.

Table 4: The case study of question recommendation.

BM25. Then, we asked 3 senior interviewers to evaluate the performance of recommendation questions. They were first required to judge which questions are relevant to this query, where the number of relevant questions is the **relevance** measure. Then, they needed to judge how many different technical aspects mentioned in those relevant questions, which is **diversity** measure, and how many questions are about **personal quality**, which should be different between technical interview (TI) and comprehensive interview (CI). As the average results shown in Table 3, we can find that compared with traditional keywords matching based approach BM25, JLMIA can recommend questions with more relevance and diversity. Meanwhile, JLMIA also can recommend more questions related to personal qualities, especially, the number of personal quality questions for comprehensive interview is more than technical interview, which distinguishes the different focuses of them two.

Further more, to illustrate the effectiveness of our question recommendation approach, we also show an example of top 4 questions recommended by different approaches in Table 4. Obviously, the given experience item is about web development. We find questions recommended by JLMIA contain all technical aspects mentioned in experience item (e.g., T1, T2 and T3 is about “*CSS and HTML programming*”, and C4 is about “*web development*”). Also, JLMIA recommends questions designed for “HTTP” (i.e., T4 and C1), which is useful knowledge for web developers. Second, for the comprehensive interview, JLMIA also recommended questions to evaluate the personal qualities of candidates, such as C2, which is related to the communication ability and problem analysis ability. Last, for the questions recommended by BM25, since they must have the same words in given requirement, the semantic relationship between keywords are neglected (e.g.,

“HTTP” and “web”). Thus, the recommended questions by BM25 do not contain more technical details.

## 6 Related Work

**Recruitment Analysis.** With the importance of talents at an all time high and the availability of recruitment big data, recruitment analysis has been attracting more and more attentions [Xu *et al.*, 2016; Zhu *et al.*, 2016]. As early as 2006, Malinowski *et al.* tried to find a good match between talents and jobs by two distinct recommendation systems [Malinowski *et al.*, 2006]. In 2011, Paparrizos *et al.* exploited all historical job transitions as well as the data associated with employees and institutions to predict the next job transition of employees [Paparrizos *et al.*, 2011]. Recently, besides the match of talents and jobs [Rác *et al.*, 2016], researchers are also devoted to analyze recruitment market from more novel perspective, such as market trend analysis [Zhu *et al.*, 2016; Lin *et al.*, 2017], career development analysis [Li *et al.*, 2017], talent circles [Xu *et al.*, 2016] and popularity measure of job skills [Xu *et al.*, 2018]. Although the above studies have explored different research aspects of recruitment market, few of them are developed for enhancing the quality and experience of job interviews. To this end, in this paper, we proposed a novel approach for intelligent job interview assessment by joint learning of multiple perspectives from large-scale real-world interview data.

**Text Mining with Topic Model.** Probabilistic topic models are capable of grouping semantic coherent words into human interpretable topics. As an important member of archetypal topic models, Latent Dirichlet Allocation (LDA) [Blei *et al.*, 2003] has a lot of extensions [Zhu *et al.*, 2014; Mimno *et al.*, 2009; Pyo *et al.*, 2015], etc.. Among them, some works focus on modeling shared latent topic distribution among multiple categories of documents, and have a wide range of practical applications. For example, Mimno *et al.* [Mimno *et al.*, 2009] designed a polylingual topic model that discovers topics aligned across multiple languages. Pyo *et al.* [Pyo *et al.*, 2015] proposed a novel model to learn the shared topic distribution between users and TV programs for TV program recommendation. Different from existing research efforts, in this paper we developed a novel model JLMIA to jointly model job description, candidate resume and interview assessment.

## 7 Concluding Remarks

In this paper, we proposed a novel approach for intelligent job interview assessment by learning the large-scale real-world interview data. To be specific, we first developed a latent variable model JLMIA to jointly model job description, candidate resume and interview assessment. JLMIA can effectively learn the representative perspectives of different job interview processes from the successful job interview records in history. Furthermore, we exploited JLMIA for two real-world applications, namely person-job fit and interview question recommendation. Extensive experiments conducted on real-world data clearly validate the effectiveness of JLMIA, which can lead to substantially less bias in job interviews and provide a valuable understanding of job interview assessment.

## Acknowledgements

This work was partially supported by grants from the National Natural Science Foundation of China (Grant No.91746301, 61727809, 61703386).

## A EM Algorithm of Variational Inference

In this appendix we give some details of the EM-style algorithm of variational inference outlined in section 3.2

First of all, we define each variational distribution term of the variational families in Equation 2. To be specific, the variational distribution of each topic proportion vector  $\theta_m^J$  is Dirichlet parameterized by vector  $\gamma_m^J$ . The variational distribution of  $\theta_{md,k}^A$ , the  $k$ -th dimension of topic proportion vector  $\theta_{md}^A$ , is univariate Gaussians  $\{\gamma_{md,k}^A, \delta^2\}$ . The variational distribution of  $z_{mj}^J$ ,  $z_{mdr}^R$  and  $z_{mde}^E$  are specified by free Multinomial with parameters  $\phi_{mj,1:K}^J$ ,  $\phi_{mdr,1:CK}^R$  and  $\phi_{mde,1:CK}^E$  respectively. The variational distribution of  $\varphi_k^J$ ,  $\varphi_k^R$ ,  $\varphi_k^{ET}$  and  $\varphi_k^{EC}$  are Dirichlet parameterized by  $\lambda_{k,1:|V^J|}^J$ ,  $\lambda_{k,1:|V^R|}^R$ ,  $\lambda_{k,1:|V^E|}^{ET}$  and  $\lambda_{k,1:|V^E|}^{EC}$ , where  $|V^J|$ ,  $|V^R|$  and  $|V^E|$  are the lengths of vocabularies of job description, resume and interview assessment, respectively.

Actually, we find each term of *ELBO* in JLMIA is similar to some parts of *ELBO* in LDA model [Blei *et al.*, 2003] or CTM model [Wang and Blei, 2011], except  $E_q[\log p(\theta_{md}^A | \theta^J, \delta^2)]$ , which can be computed by:

$$\begin{aligned} E_q[\log p(\theta_{md}^A | \theta^J, \delta^2)] &= E_q[\log N(\theta_{md}^A | h(\theta_m^J, C), \delta^2 I)] = \\ &= -\frac{CK}{2}(\log \delta^2 + \log 2\pi) - \frac{1}{2\delta^2} \sum_{k=1}^{CK} E_q[(\theta_{md,k}^A - \log \theta_{m,k'}^J)^2], \\ E_q[(\theta_{md,k}^A - \log \theta_{m,k'}^J)^2] &= \delta^2 + \Psi'(\gamma_{m,k'}^J) - \Psi'(|\gamma_{m,1:K}^J|) \\ &\quad + (\gamma_{md,k}^A - \Psi(\gamma_{m,k'}^J) + \Psi(|\gamma_{m,1:K}^J|))^2, \end{aligned}$$

where we assume that  $|\gamma_{m,1:K}^J| = \sum_{i=1}^K \gamma_{m,i}^J$ , and  $k' = k \bmod K$ . Similar symbols are not described later for simplicity. And the  $\Psi(\cdot)$  is Digamma function with derivative  $\Psi'(\cdot)$ .

Then, we describe our EM-style algorithm. In E-step, we employ coordinate ascent approach to optimize all variational parameters. First, we optimize the  $\zeta_{md}$  in Equation 4:

$$\hat{\zeta}_{md} = \sum_{k=1}^{CK} \exp\{\gamma_{md,k}^A + \delta^2/2\}.$$

Second, we optimize  $\phi_{mj,1:K}^J$ ,  $\phi_{mdr,1:CK}^R$  and  $\phi_{mde,1:CK}^E$  for each coordinate. Assume that  $w_{mj}^J = c$ ,  $w_{mdr}^R = t$  and  $w_{mde}^E = i$ ,  $I_{mde} = TI$ :

$$\begin{aligned} \hat{\phi}_{mj,k}^J &\propto \exp\{\Psi(\lambda_{k,c}^J) - \Psi(|\lambda_{k,1:|V^J|}^J|) + \Psi(\gamma_{m,k}^J) - \Psi(|\gamma_{m,1:K}^J|)\}, \\ \hat{\phi}_{mdr,k}^R &\propto \exp\{\Psi(\lambda_{k,t}^R) - \Psi(|\lambda_{k,1:|V^R|}^R|) + \gamma_{md,k}^R\}, \\ \hat{\phi}_{mde,k}^E &\propto \exp\{\Psi(\lambda_{k,i}^{ET}) - \Psi(|\lambda_{k,1:|V^E|}^{ET}|) + \gamma_{md,k}^E\}. \end{aligned}$$

Third, we optimize  $\gamma_m^J$ . Due to no analytic solution, we use Newton's method for each coordinate:

$$\begin{aligned} \frac{dELBO}{d\gamma_{m,i}^J} &= -\frac{1}{2\delta^2} \sum_{d=1}^{D_m} \sum_{k=1}^{CK} \left( 2(\Psi(\gamma_{m,k'}^J) - \Psi(|\gamma_{m,1:K}^J|) - \gamma_{md,k}^A) \right. \\ &\quad \times (\delta_{k'}^i \Psi'(\gamma_{m,k'}^J) - \Psi'(|\gamma_{m,1:K}^J|)) + \delta_{k'}^i \Psi''(\gamma_{m,k'}^J) - \Psi''(|\gamma_{m,1:K}^J|) \Big) \\ &\quad + \sum_{k=1}^K (|\phi_{m1:N_m^J,k}^J| + \alpha_k - \gamma_{m,k}^J) (\delta_k^i \Psi'(\gamma_{m,k}^J) - \Psi'(|\gamma_{m,1:K}^J|)), \end{aligned}$$

where function  $\delta_x^y = 1$ , only if  $x = y$ , otherwise,  $\delta_x^y = 0$ .

Fourth, we optimize  $\gamma_{m,d,1:CK}^A$ . Due to no analytic solution, again, we use conjugate gradient algorithm with derivative:

$$\frac{dELBO}{d\gamma_{m,d,k}^A} = -\frac{1}{\delta^2} \left( \gamma_{m,d,k}^A - \Psi(\gamma_{m,k}^J) + \Psi(|\gamma_{m,1:k}^J|) \right) + |\phi_{m,d,1:N_{m,d}^R}^R| + |\phi_{m,d,1:N_{m,d}^E}^E| - (N_{m,d}^R + N_{m,d}^E) \zeta_{m,d}^{-1} \exp\{\gamma_{m,d,k}^A + \delta^2/2\}.$$

Last, we optimize  $\lambda^J$ ,  $\lambda^R$ ,  $\lambda^{ET}$  and  $\lambda^{EC}$ . Their calculation process are similar, token  $\lambda_{k,c}^J$  and  $\lambda_{k,i}^{ET}$  as examples:

$$\lambda_{k,c}^J = \beta_c^J + \sum_{m=1}^M \sum_{j=1}^{N_m^J} \phi_{m,j,k}^J \delta_{w_j}^c,$$

$$\lambda_{k,i}^{ET} = \beta_i^E + \sum_{m=1}^M \sum_{d=1}^{D_m} \sum_{e=1}^{N_{m,d}^E} \phi_{m,d,e,k}^J \delta_{w_{mde}}^i \delta_{I_{mde}}^{TI}.$$

In the M-step, we maximize the *ELBO* with respect to parameter  $\alpha$ , similar to LDA, and regard the other hyperparameters in  $\Omega$  as fixed parameters.

## References

- [Bersin, 2013] Josh Bersin. <https://www.forbes.com/sites/joshbersin/2013/05/23/corporate-recruitment-transformed-new-breed-of-service-providers/>. 2013.
- [Blei *et al.*, 2003] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.
- [Blei *et al.*, 2017] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, (just-accepted), 2017.
- [Li *et al.*, 2017] Huayu Li, Yong Ge, Hengshu Zhu, Hui Xiong, and Hongke Zhao. Prospecting the career development of talents: A survival analysis perspective. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*, pages 917–925, 2017.
- [Lin *et al.*, 2017] Hao Lin, Hengshu Zhu, Yuan Zuo, Chen Zhu, Junjie Wu, and Hui Xiong. Collaborative company profiling: Insights from an employee’s perspective. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 1417–1423, 2017.
- [Malinowski *et al.*, 2006] Jochen Malinowski, Tobias Keim, Oliver Wendt, and Tim Weitzel. Matching people and jobs: A bilateral recommendation approach. In *System Sciences, 2006. HICSS’06. Proceedings of the 39th Annual Hawaii International Conference on*, volume 6, pages 137c–137c. IEEE, 2006.
- [Mimno *et al.*, 2009] David Mimno, Hanna M Wallach, Jason Naradowsky, David A Smith, and Andrew McCallum. Polylingual topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 880–889. Association for Computational Linguistics, 2009.
- [Paparrizos *et al.*, 2011] Ioannis Paparrizos, B Barla Cambazoglu, and Aristides Gionis. Machine learned job recommendation. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 325–328. ACM, 2011.
- [Pyo *et al.*, 2015] Shinjee Pyo, Eunhui Kim, et al. Lda-based unified topic modeling for similar tv user grouping and tv program recommendation. *IEEE transactions on cybernetics*, 45(8):1476–1490, 2015.
- [Rácz *et al.*, 2016] Gábor Rácz, Attila Sali, and Klaus Dieter Schewe. *Semantic Matching Strategies for Job Recruitment: A Comparison of New and Known Approaches*. Springer International Publishing, 2016.
- [Tang *et al.*, 2014] Fangshuang Tang, Qi Liu, Hengshu Zhu, Enhong Chen, and Feida Zhu. Diversified social influence maximization. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 455–459. IEEE, 2014.
- [Wang and Blei, 2011] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’11*, pages 448–456, New York, NY, USA, 2011. ACM.
- [Wang and McCallum, 2006] Xuerui Wang and Andrew McCallum. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433. ACM, 2006.
- [Xu *et al.*, 2016] Huang Xu, Zhiwen Yu, Jingyuan Yang, Hui Xiong, and Hengshu Zhu. Talent circle detection in job transition networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 655–664. ACM, 2016.
- [Xu *et al.*, 2018] Tong Xu, Hengshu Zhu, Chen Zhu, Pan Li, and Hui Xiong. Measuring the popularity of job skills in recruitment market: A multi-criteria approach. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, February 2-7, 2018, New Orleans, Louisiana, USA.*, 2018.
- [Zhang *et al.*, 2014] Yingya Zhang, Cheng Yang, and Zhixiang Niu. A research of job recommendation system based on collaborative filtering. In *Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium on*, volume 1, pages 533–538. IEEE, 2014.
- [Zhu *et al.*, 2014] Chen Zhu, Hengshu Zhu, Yong Ge, Enhong Chen, and Qi Liu. Tracking the evolution of social emotions: A time-aware topic modeling perspective. In *2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, December 14-17, 2014*, pages 697–706, 2014.
- [Zhu *et al.*, 2016] Chen Zhu, Hengshu Zhu, Hui Xiong, Pengliang Ding, and Fang Xie. Recruitment market trend analysis with sequential latent variable models. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 383–392. ACM, 2016.