A Joint Learning Approach to Intelligent Job Interview Assessment

Dazhong Shen1,2, Hengshu Zhu2,*, Chen Zhu2, Tong Xu1,2, Chao Ma2, Hui Xiong1,2,3,4,*

1 Anhui Province Key Lab of Big Data Analysis and Application, University of S&T of China,
2 Baidu Talent Intelligence Center,
3 Business Intelligence Lab, Baidu Research,
4 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.
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., $J_m = \{w_{mj}\}_{j=1}^{N_m}$, 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 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 $\theta^A_{md}$ over topics. Second, for revealing the relationships between job descriptions and resumes along with the differences between their diversity, we generate $\theta^A_{md}$ from the logistic-normal distribution with mean parameter related to the topic distribution of job description $\theta^j_{md}$. And the topic numbers of $\varphi^j$, $\varphi^R$ and $\varphi^E$ are set as $|J| = |R| = C \cdot |J| = CK$. 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., Technical Interview or Comprehensive Interview) 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., $h(\theta^j_{md}, C) = \log \theta^j_{md,k'}$, $k' = k \ mod \ K$, $1 \leq k' \leq K$, and $\pi(\theta)$, in line 3.(b).i and 3.(c).i, is the logistic transformation, i.e., $\pi(\theta^j_{md,k}) = \frac{\exp(\theta^j_{md,k})}{\sum_{i=1}^{K} \exp(\theta^j_{md,i})}$.

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.

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

1. For each topic $k$ of candidate interview record:
   a. Draw $\varphi^j_k$ from the Dirichlet prior $\text{Dir}(\beta^j)$.
   b. Draw $\varphi^E_k$ and $\varphi^E_k$ from the Dirichlet prior $\text{Dir}(\beta^j)$.
2. For each job description $J_m$:
   a. Sample topic distribution $\theta^j_{md} \sim \text{Dir}(\alpha)$.
   b. For each candidate interview record pair $(R_{md}, E_{md}, I_{md})$:
      i. Draw topic assignment $z^R_{md} \sim \text{Multi}(\pi(\theta^A_{md}))$.
      ii. Draw word $w^R_{md} \sim \text{Multi}(\varphi^j_{z^R_{md}})$.
      iii. For the $c$-th word $w^E_{md}$ in interview assessment $E_{md}$:
          i. Draw topic assignment $z^E_{md} \sim \text{Multi}(\pi(\theta^A_{md}))$.
          ii. Draw word $w^E_{md} \sim \text{Multi}(\varphi^{EC}_{z^E_{md}})$.
   iii. Draw word $w^E_{md} \sim \text{Multi}(\varphi^{ET}_{z^E_{md}})$ ($I_{md} == TT$).
   iv. Draw word $w^E_{md} \sim \text{Multi}(\varphi^{EC}_{z^E_{md}})$ ($I_{md} == CT$).
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 Φ and all hyper-parameters by Ω. Following the generative process, the joint distribution can be factored as:

\[ p(S, Φ|Ω) = p(Φ|Ω) \prod_{m=1}^{M} P(S_m|Φ), \]

where each component can be calculated by:

\[ p(S_m|Φ) = p(J_m|z^J_m, ϕ^J_m) \prod_{d=1}^{D_m} p(R_m|z^R_m, ϕ^R_m)p(E_m|z^E_m, ϕ^E_m, I_m), \]

\[ p(Φ|Ω) = \prod_{m=1}^{M} \prod_{d=1}^{D_m} p(ϕ^{A,m}_d, θ^J_m, S^J_m) \prod_{r=1}^{R_m} p(ϕ^{R,m}_r, ϕ^{A,m}_d) \prod_{e=1}^{E_m} p(ϕ^{E,m}_e, θ^A_m). \]

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:

\[ q(Φ) = \prod_{m=1}^{M} q(ϕ^{A,m}_d) \prod_{r=1}^{R_m} q(ϕ^{R,m}_r)q(ϕ^{E,m}_e) \prod_{m=1}^{M} q(θ^J_m) \prod_{m=1}^{M} q(z^J_m), \]

\[ \times \prod_{m=1}^{M} \prod_{d=1}^{D_m} q(ϕ^{A,m}_d, θ^J_m) \prod_{r=1}^{R_m} q(ϕ^{R,m}_r, θ^J_m) \prod_{e=1}^{E_m} q(ϕ^{E,m}_e, θ^A_m). \]

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):

\[ \log p(S|Ω) \geq E_q[\log p(S|Ω)] + H(q) \]

\[ = E_q[\log p(Φ|Ω)] + \sum_{m=1}^{M} E_q[\log p(S_m|Φ)] + H(q), \]

where the expectation \( E_q[.] \) 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 excepted log probability of topic assignments in documents of each candidate interview records. Similar to [Wang and Blei, 2011], we introduce a new variational parameter \( z \in \{z_{m, r} \mid D_m \}_{m=1}^{M} \) to preserve the lower bound of ELBO. Here we take the \( E_q[\log p(z_{m, r}|ϕ^A)] \) as an example to explain it (the \( E_q[\log p(z_{m, r}|ϕ^A)] \) can be computed in a similar way):

\[ E_q[\log p(z_{m, r}|ϕ^A)] = E_q[ϕ^A | z_{m, r}] - E_q[\log \sum_{k=1}^{K} \exp(ϕ^{A,m}_d)]. \]

\[ \geq E_q[ϕ^{A,m}_d | z_{m, r}] - C^{CK}_{m, r} \sum_{k=1}^{K} E_q[\exp(ϕ^{A,m}_d)] + \log(\sum_{k=1}^{K} \exp(ϕ^{A,m}_d)). \]

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, \( ϕ^J_g \) and \( ϕ^R_g \), are learned, we can compute the document-topic distribution by:

\[ \theta^J_{g, k} = \frac{1}{N^J} \sum_{m=1}^{N^J} ϕ^J_g, k = 1, \cdots, K. \]

\[ \theta^R_{g, k} = \frac{1}{N^R} \sum_{m=1}^{N^R} ϕ^R_g, k = 1, \cdots, K. \]

Then, by computing the similarity between \( \theta^J_g \) and \( \theta^R_g \), 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 \( \theta^J_g \) and \( \theta^R_g \) may be different, here we have:

\[ \tilde{θ}^R_g = \sum_{c \in C_k} θ^R_g, k = 1, \cdots, K, \]

where \( C_k \) is a set of mapping index that satisfies \( θ^R_g, c \in C_k \) is generated from \( θ^J_{g, k} \). In particular, the vectors \( θ^J_g \) and \( θ^R_g \) 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 \( T \) (i.e., job requirement item or experience item of candidate). Similar to the process of computing the topic distributions \( θ^R_g \), we can compute the topic distribution \( θ^Q_g \) of each question \( q_i \in Q \), through regarding interview questions as a part of interview assessment. Let \( θ^Q_g \) denote a latent representation of question \( q_i \) and the latent representation of the given query \( T \) as \( θ^T_g \in \{θ^J_g, θ^R_g\} \).

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 \( θ^T_g \), 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:

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

$$= \mu \sum_{i=1}^{\text{dim}} \text{Sim}(q_i, \theta_i^j) + (1 - \mu) \sum_{i=1}^{\text{dim}} \text{Div}(\theta_i^j)$$

s.t. $X \subset Q$, $|X| = L$, $0 < \mu < 1$, (6)

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

In general, the calculation of addressing $F(X)$ is computationally prohibitive, since we will assemble the explosion problem if we calculate Equation 6 for all subsets $X$. Fortunately, since $F(X)$ defined in this paper is submodular [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 $\{\alpha, \beta, \beta^R, \beta^L\} = \{0.01, 0.1, 0.1, 0.1\}$. 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 can observe 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 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.

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Resume</th>
<th>Tech. Interview</th>
<th>Com. Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>Topic 1</td>
<td>Topic 1</td>
<td>Topic 1</td>
</tr>
<tr>
<td>Experience</td>
<td>Function</td>
<td>Foundation</td>
<td>Technology</td>
</tr>
<tr>
<td>Foundation</td>
<td>Management</td>
<td>Knowledge</td>
<td>Communication</td>
</tr>
<tr>
<td>Technology</td>
<td>Backstage</td>
<td>Code</td>
<td>Study</td>
</tr>
<tr>
<td>PHP</td>
<td>HTTP</td>
<td>Development</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Web</td>
<td>Module</td>
<td>Development</td>
<td>Development</td>
</tr>
<tr>
<td>Proficient</td>
<td>Topic 1</td>
<td>Topic 1</td>
<td>Topic 1</td>
</tr>
<tr>
<td>Engineer</td>
<td>Web Site</td>
<td>JS</td>
<td>Job</td>
</tr>
<tr>
<td>Interest</td>
<td>Web Page</td>
<td>Methods</td>
<td>Pressure</td>
</tr>
<tr>
<td>Development</td>
<td>System</td>
<td>CSS</td>
<td>Solution</td>
</tr>
<tr>
<td>Web Page</td>
<td>Web</td>
<td>Elements</td>
<td>Like</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Framework</td>
<td>Events</td>
<td>Work Overtime</td>
</tr>
</tbody>
</table>

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

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

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 $R\ell = 5$, $Div = 20$ and $\mu = 0.9$.

We randomly selected 100 experience items as the queries. For each query, we recommend 10 questions by JLMIA and 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.,
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 [Rácz 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 devoted 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 \( \theta^{A}_{m,j,k} \) is Dirichlet parameterized by vector \( \gamma^{A}_{m,j,k} \). The variational distribution of \( \gamma^{A}_{m,j,k} \), the \( k \)-th dimension of topic proportion vector \( \theta^{A}_{m,j,k} \), is univariate Gaussians \( \gamma^{A}_{m,j,k} \). The variational distribution of \( \zeta^{J}_{m,j} \), \( \zeta^{R}_{mdr} \), and \( \zeta^{E}_{mde} \) are specified by free Multi-nomial with parameters \( \phi^{J}_{m,j,1:K} \), \( \phi^{R}_{mdr,1:CK} \), and \( \phi^{E}_{mde,1:CK} \) respectively. The variational distribution of \( \varphi^{F}_{k,t} \), \( \varphi^{R}_{k,t} \), \( \varphi^{E}_{k,1:|V^{F}_{j}|} \), \( \varphi^{F}_{k,1:|V^{R}_{j}|} \), \( \varphi^{E}_{k,1:|V^{E}_{j}|} \), \( \lambda^{F}_{k,1:|V^{F}_{j}|} \), \( \lambda^{R}_{k,1:|V^{R}_{j}|} \), and \( \lambda^{E}_{k,1:|V^{E}_{j}|} \), where \( |V^{F}_{j}| \), \( |V^{R}_{j}| \), and \( |V^{E}_{j}| \) 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 \( E_q[\log(\theta^{A}_{m,j,k} | \vartheta^{A}_{m,j,k}, \delta^{2})] \), which can be computed by:

\[
E_q[\log p(\theta^{A}_{m,j,k} | \vartheta^{A}_{m,j,k}, \delta^{2})] = E_q[\log N(\theta^{A}_{m,j,k} | \langle \theta^{A}_{m,j,k}, \vartheta^{A}_{m,j,k} \rangle, \delta^{2}I)] = -CK(\log \delta^{2} + 2\log \pi - \frac{1}{2\delta^{2}} \sum_{k=1}^{K} \sum_{i=1}^{C} E_q[\langle \theta^{A}_{m,j,k} - \log \theta^{A}_{m,j,k} \rangle^2])
\]

where we assume that \( |\gamma^{J}_{m,j,1:K}| = \sum_{k=1}^{K} \sum_{j=1}^{J} \gamma^{J}_{m,j,k} \), and \( k^{'} = k \mod K \). 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 \( \zeta^{md}_{m,j} \) in Equation 4:

\[
\xi^{md}_{m,j} = \frac{CK}{\delta^{2}} \sum_{k=1}^{K} \exp(\theta^{J}_{m,j,k} + \delta^{2}/2)
\]

Second, we optimize \( \phi^{J}_{m,j,1:K} \), \( \phi^{R}_{mdr,1:CK} \), and \( \phi^{E}_{mde,1:CK} \) for each coordinate. Assume that \( w^{l}_{m,j} = c, w^{R}_{m,j} = t \) and \( w^{E}_{mde} = i, l^{md}_{m,j} = Tl^{T} \):

\[
\begin{align*}
\phi^{J}_{m,j,k} &\propto \exp(\Psi(\lambda^{J}_{m,j,k}')) - \Psi(\lambda^{J}_{m,j,k}')) + \Psi(\gamma^{J}_{m,j,k}') - \Psi(\gamma^{J}_{m,j,k}')), \\
\phi^{R}_{mdr,k} &\propto \exp(\Psi(\lambda^{R}_{mdr,k})) - \Psi(\lambda^{R}_{mdr,k})) + \Psi(\gamma^{R}_{m,j,k}')) - \Psi(\gamma^{R}_{m,j,k}')), \\
\phi^{E}_{mde,k} &\propto \exp(\Psi(\lambda^{E}_{mde,k})) - \Psi(\lambda^{E}_{mde,k})) + \Psi(\gamma^{E}_{mde,k}')) - \Psi(\gamma^{E}_{mde,k}')).
\end{align*}
\]

Third, we optimize \( \gamma^{J}_{m,j} \). Due to no analytic solution, we use Newton’s method for each coordinate:

\[
\begin{align*}
dELBO_{d}\gamma^{J}_{m,j,k} = -\frac{1}{2\delta^{2}} \sum_{d=1}^{Dm} \sum_{j=k}^{K} (2(\Psi(\gamma^{J}_{m,j,k})') - \Psi(\gamma^{J}_{m,j,k}'')) - \Psi(\gamma^{J}_{m,j,k}'))
\end{align*}
\]

Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

3547
where function $\delta_y = 1$, only if $x = y$, otherwise, $\delta_y = 0$.

Fourth, we optimize $A_{md;1:CK}$. Due to no analytic solution, again, we use conjugate gradient algorithm with derivative:

\[
\frac{\partial \text{ELBO}}{\partial \gamma^A_{md,k}} = -\frac{1}{\alpha} \left( \gamma^A_{md,k} - \psi(\gamma^A_{md,k}) + \psi'((\gamma^A_{md,k} + \beta^A_{md,k})) + \delta^A_{md;1:NR_{md},k} \right) \\
+ \varepsilon^E_{md;1:NR_{md},k} - (N^R_{md} + N^E_{md}) \gamma^A_{md,k} + \delta^2 / 2. 
\]

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

\[
\lambda^J_{k,c} = \beta^J_c + \sum_{m=1}^{M} \sum_{j=1}^{N^J_m} \phi^J_{m,k,j} \delta^J_{m,j} \\
\lambda^{ET}_{k,i} = \beta^E_t + \sum_{m=1}^{M} \sum_{d=1}^{D_m} \sum_{e=1}^{N^E_{md}} \phi^{ET}_{mde,k} \delta^{ET}_{mde}.
\]

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

References


