Enhancing Employer Brand Evaluation with Collaborative Topic Regression Models

HAO LIN, School of Economics and Management, Beihang University, China
HENGSHU ZHU, Talent Intelligence Center, Baidu Inc., China
JUNJIE WU, School of Economics and Management, Beihang University, China and Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, China
YUAN ZUO, School of Economics and Management, Beihang University
CHEN ZHU, Talent Intelligence Center, Baidu Inc., China
HUI XIONG, Management Science and Information Systems Department, Rutgers University, USA

Employer Brand Evaluation (EBE) is to understand an employer's unique characteristics to identify competitive edges. Traditional approaches rely heavily on employers' financial information, including financial reports and filings submitted to the Securities and Exchange Commission (SEC), which may not be readily available for private companies. Fortunately, online recruitment services provide a variety of employers' information from their employees' online ratings and comments, which enables EBE from an employee's perspective. To this end, in this article, we propose a method named Company Profiling–based Collaborative Topic Regression (CPCTR) to collaboratively model both textual (i.e., reviews) and numerical information (i.e., salaries and ratings) for learning latent structural patterns of employer brands. With identified patterns, we can effectively conduct both qualitative opinion analysis and quantitative salary benchmarking. Moreover, a Gaussian processes–based extension, GPCTR, is proposed to capture the complex correlation among heterogeneous information. Extensive experiments are conducted on three real-world datasets to validate the effectiveness and generalizability of our methods in real-life applications. The results clearly show that our methods outperform state-of-the-art baselines and enable a comprehensive understanding of EBE.

CCS Concepts: • **Information systems → Data mining**; • **Computing methodologies → Machine learning**; • **Applied computing → Business intelligence**;

A preliminary version of this article has been published in the 31st AAAI Conference on Artificial Intelligence (AAAI-17) [Lin et al. 2017].
Dr. Junjie Wu's work was partially supported by the National Key R&D Program of China (2019YFB2101804), the National Special Program on Innovation Methodologies (SQ2019IM4910001), and the National Natural Science Foundation of China (71725002, 71531001, U1636210). Dr. Yuan Zuo was supported by the National Natural Science Foundation of China under Grant 71901012 and the China Postdoctoral Science Foundation under Grant 2018M640045.
Authors' addresses: H. Lin and Y. Zuo, School of Economics and Management, Beihang University, No. 37 Xue Yuan Road, Haidian District, Beijing, 100191, China; emails: {linhao2014, zuoyuan}@buaa.edu.cn; H. Zhu (corresponding author) and C. Zhu, Talent Intelligence Center, Baidu Inc., 10 Shangdi 10th Street, Haidian District, Beijing, 100085, China; emails: zhuhengshu@gmail.com, zhuchen02@baidu.com; J. Wu (corresponding author), School of Economics and Management, Beihang University, No. 37 Xue Yuan Road, Haidian District, Beijing, 100191, China, and Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, No. 37 Xue Yuan Road, Haidian District, Beijing, 100191, China; email: wujj@buaa.edu.cn; H. Xiong, Management Science and Information Systems Department, Rutgers University, Newark, NJ, 07102, USA; email: hxiong@rutgers.edu.

1046-8188/2020/05-ART32 $15.00
https://doi.org/10.1145/3392734

Additional Key Words and Phrases: Employer brand evaluation, collaborative topic regression, salary benchmarking, Gaussian processes

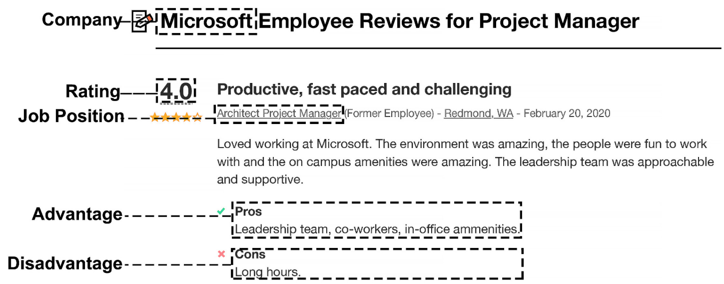
ACM Reference format:
Hao Lin, Hengshu Zhu, Junjie Wu, Yuan Zuo, Chen Zhu, and Hui Xiong. 2020. Enhancing Employer Brand Evaluation with Collaborative Topic Regression Models. *ACM Trans. Inf. Syst.* 38, 4, Article 32 (May 2020), 33 pages.
https://doi.org/10.1145/3392734

1 INTRODUCTION

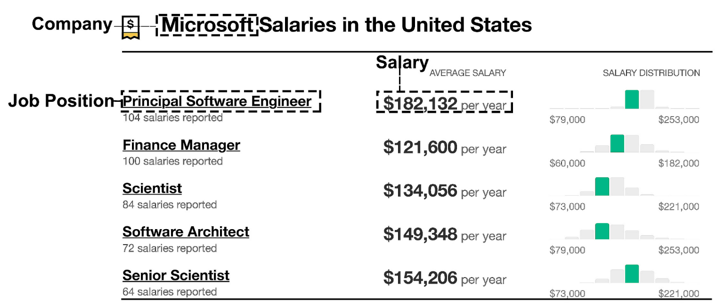
Brands are one of the most precious assets for a company, and it is crucial for organizations to manage brands as a strategic tool to keep up with the continuously changing business world. The employer brand, which highlights an organization's corporate image attributes (e.g., attractive products and services), people-and-culture attributes (e.g., commitment to diversity and inclusion), job characteristics (e.g., flexible working conditions), or other unique aspects of the organization as an employer, is raising increasing attention in the area of human resource management. Employers build superior brand images by developing directional strategies to manage people's consciousness, which is termed as "employer branding" [Backhaus and Tikoo 2004; Chhabra and Sharma 2014; Edwards 2009; Foster et al. 2010; Ito et al. 2013; Xie et al. 2015]. Moreover, with the employer brand image, applicants can match their job pursuit intentions with the traits of a candidate company. To this end, it is of great importance to build effective strategies for evaluating the brand competitiveness of an employer, which is an analytical process of understanding the employer's unique characteristics that helps to distinguish the target company as a competitive employer from the others.

In the past decades, traditional approaches for Employer Brand Evaluation (EBE) have relied heavily on the availability of the rich finance information about the employer, such as finance reports and filings submitted to the Securities and Exchange Commission (SEC). However, due to data privacy issues, this finance information may not be readily available for many employers, especially the private companies. Recently, with the rapid prevalence of online employment services (e.g., Glassdoor, Indeed, and Kanzhun), a new research paradigm is enabled for obtaining a large amount of employers' information, including varieties of behavioral data anonymously published by (former) employees of a company, such as reviews, salaries, Q/As, interviews, and so on. Among this public information, reviews (see Figure 1(a)), and salaries (see Figure 1(b)) may play key roles in enhancing EBE, since they are both strong indicators of employers' attractiveness/competitiveness under an employment environment and may greatly influence the employment decisions of (potential) applicants. For example, reviews may reveal employees' perceptions of the job (e.g., management environment in Figure 1(a)), which are shown to contribute as significant predictors of employer brand in several previous studies [Agrawal and Swaroop 2009; Lievens and Highhouse 2003; Lievens et al. 2007]. The effect of salary to employer brand attractiveness is also validated by existing research works [Agrawal and Swaroop 2009; Berthon et al. 2005; Lievens and Highhouse 2003; Lievens et al. 2007]. Nevertheless, none of existing works provide data-driven methods for EBE.

Motivated by the above issues, we attempt to simultaneously utilize reviews (i.e., ratings and comments) and salaries to evaluate employer brand in a fully automated manner; that is, to not only help employers to qualitatively identify their business advantages and disadvantages based on reviews (namely, opinion analysis), but also to quantitatively benchmark their market competitiveness through salaries with respect to different job positions (also called salary benchmarking [Meng et al. 2018]). For salary benchmarking, we construct a job-company salary matrix and



(a) Review page for Microsoft.



(b) Salary profile page for Microsoft.

Fig. 1. An example of Indeed pages.

aim to predict missing entries in the matrix. The severe sparsity in the salary matrix poses great challenges to salary benchmarking. To alleviate the sparsity problem, we treat review ratings and comments as complementary of each other, both of which are in turn regarded as side information for salary benchmarking. Therefore, in this article, we need to model both the textual indicator (i.e., review texts) and numerical indicators (i.e., salaries and ratings) of the corresponding job-company pair collaboratively.

Despite their indication of employer brand attractiveness, reviews and salaries are typical heterogeneous information, which require fusion model to seamlessly combine all this information for EBE. To bridge different information, we turn to the research paradigm of Collaborative Topic Regression (CTR) [Wang and Blei 2011], a typical joint method to model both textual and numerical information. More specifically, we propose a model named Company Profiling–based Collaborative Topic Regression (CPCTR) to formulate a joint probabilistic graphical model for learning the latent patterns of employer brands, which is among the earliest to incorporate heterogeneous information for EBE. Considering that review texts contain both pros and cons textual segments, we introduce a probabilistic topic model for capturing both positive and negative opinion patterns of a job-company pair. For the generation of numerical information such as ratings and salaries, a latent factor model based on probabilistic matrix factorization is chosen to factorize the observed matrix into latent factors of jobs and job-company pairs. Further, a latent offset is introduced to bridge the gap between factorized job-company latent variable and topic proportions learned by topic modeling.

Moreover, real-world heterogeneous information usually shows complex correlation. For example, in our settings, salary of a job-company pair may have non-linear correlation with its corresponding review text or rating. In other words, a high-salary employee might feel unsatisfied with his working conditions, driving him to rate his job with a relatively low score and publish more negative comments towards his employer. To that end, an enhanced model named Gaussian Processes–based Collaborative Topic Regression (GPCTR) is proposed to capture complex correlation among heterogeneous review texts, ratings, and salaries. GPCTR utilizes Gaussian processes for complex correlation modeling, since Gaussian processes uses flexible covariance functions and can therefore better model complicated mappings between latent variables and observed values. Additionally, Gaussian processes is a typical non-parametric Bayesian model and can easily be extended with our existing probabilistic graphical model framework.

Meanwhile, we conduct extensive experiments on real-world datasets for validating the performance of CPCTR and GPCTR. Specifically, we design salary prediction experiments to demonstrate that both CPCTR and GPCTR can achieve more effective prediction performance compared to state-of-the-art baselines. Contribution of each data source is also investigated to show that review texts and ratings are both essential for salary prediction. Particularly, empirical experiments show that GPCTR can better capture the complex correlation across different data sources while CPCTR provides a comprehensive interpretation of employer brand characteristics. Besides, by evaluating the outputs learned by CPCTR, it is interesting to observe many meaningful patterns of employer brands, such as *welfare and technology are the typical pros of Baidu, while those of Tencent are training and learning*. Finally, we conduct a price prediction experiment for demonstrating that our proposed models are generally applicable over areas other than EBE, such as product brand evaluation.

This article is a substantial extension of our previous conference version [Lin et al. 2017]. Compared to the preliminary work [Lin et al. 2017], this work focuses on a more realistic application domain, namely data-driven EBE, which opens a brand new research direction. As for methodology, this work extends the original CPCTR model to an enhanced model named GPCTR, under the framework of fusing heterogeneous review texts, ratings and salaries for EBE. Specifically, this work believes that the heterogeneity of different information involves with complex correlation that can be better captured by Gaussian processes whereas the conference version [Lin et al. 2017] only models it with a Gaussian latent offset variable. Furthermore, this work validates the robustness of the proposed methods and their generalizability in other application setting on two more datasets, including Glassdoor and Amazon, while previous work [Lin et al. 2017] only conducts experiments on Kanzhun.

The remainder of this article is arranged as follows: Section 2 introduces the related work of this article. In Section 3, data description and problem settings are given. Section 4 shows the technical details of our proposed models, including both CPCTR and GPCTR. We perform extensive empirical studies in Section 5 and conclude the work in Section 6.

2 RELATED WORK

The related work of this article can be broadly categorized into two facets: (1) application-driven problem contexts including *employer branding*, *data-driven talent analytics,* and *salary prediction*; (2) more general machine learning techniques including *topic modeling for opinion analysis* and *matrix factorization for prediction.*

2.1 Employer Branding

Employer brand refers to the reputation of an employer as a place to work and employees' perceptions towards the employer, which differs from the more general company brand. By building and enhancing employer brand, employers aim to attract and retain talented employees, which is termed as *employer branding* and has long been a research focus in the community of human resource management and marketing disciplines. Backhaus and Tikoo [2004] developed a conceptual framework for understanding employer branding and indicated that employer branding helps to affect employer attractiveness and shape employer loyalty. Many other studies have attempted to identify the key components of employer attractiveness or employer image in employer branding [Agrawal and Swaroop 2009; Berthon et al. 2005; Lievens and Highhouse 2003; Lievens et al. 2007]. For example, Berthon et al. [2005] proposed to use factor analysis to find significant factors that influence employer attractiveness.

This line of research indeed motivates us to perform EBE by benchmarking employer competitiveness through review texts and salaries. This is because review texts contain many words reflecting employees' perceptions of the job (e.g., working conditions, social support), which are shown to contribute as significant predictors of employer brand in several previous studies [Agrawal and Swaroop 2009; Lievens and Highhouse 2003; Lievens et al. 2007]. The effect of salary to indicate employer brand attractiveness is also proved by existing research works [Agrawal and Swaroop 2009; Berthon et al. 2005; Lievens and Highhouse 2003; Lievens et al. 2007].

2.2 Data-driven Talent Analytics

Talent analytics [Mohapatra and Sahu 2018; NCho 2017] aims at employing advanced data-driven techniques for conducting talent management and has become a research focal point in application-driven business intelligence for enterprises. Recently, with prevalence of big talent data, various novel research perspectives are enabled, including person-job fit [Qin et al. 2018; Yan et al. 2019; Zhu et al. 2018], person-organization fit [Sun et al. 2019], candidate recommendation for talent search [Geyik et al. 2019, 2018a, 2018b; Ha-Thuc et al. 2017; Ozcaglar et al. 2019; Ramanath et al. 2018], and talent flow analysis [Xu et al. 2018; Zhang et al. 2019]. For instance, Zhu et al. [2018] proposed to learn representations from both job postings and resumes for person-job fit prediction. Qin et al. [2018, 2020] utilized the rich historical job application information and proposed a neural network approach with hierarchical ability-aware attention for person-job fit modeling. To tackle the problem of insufficient labeled data for person-job matching learning, Bian et al. [2019] built a domain adaptation method to utilize rich information of a source domain for improved prediction in a target domain with limited labeled data. Sun et al. [2019] proposed to measure the compatibility between employees and their corresponding employers. Ha-Thuc et al. [2017] leveraged large quantities of semi-structured data and social affiliations on LinkedIn for building a Query-by-Example talent search system. For large-scale talent flow forecast, Zhang et al. [2019] designed a dynamic latent factor method, which jointly models historical talent flows, market influence, and company attributes.

Different from the above studies, we attempt to utilize the typical kinds of talent data, including reviews and salaries extracted from online recruitment services, for tackling a brand new research topic, namely, EBE. We believe that data-driven EBE would be of great interest to both academia and industry.

2.3 Salary Prediction

Salary prediction is an important research problem due to the privacy nature of salary information in the business environment. Traditional approaches for salary prediction usually either rely heavily on limited survey data or employ regression-based machine learning models. For example, Lazar [2004] used support vector machine on survey data provided by the US Census Bureau for income prediction. Khongchai and Songmuang [2016] applied five methods including decision trees, K-nearest neighbor, neural networks, support vector machines, and naive Bayes for predicting salaries of individual students. Recently, LinkedIn salary prediction is raising increasing attention [Chen et al. 2018; Kenthapadi et al. 2017]. Kenthapadi et al. [2017] proposed a Bayesian hierarchical smoothing–based method for computation of robust compensation insights in LinkedIn.

Chen et al. [2018] utilized rich information in LinkedIn Economic Graph for computing LinkedIn salary, specifically by learning semantic representations from company transition data.

A closely related work is Meng et al. [2018], where a matrix factorization method is proposed for salary prediction by integrating multiple confounding factors such as company or job similarity. However, this method requires extra computation of similarity matrix, which needs auxiliary company information or job descriptions. Our proposed models are quite different from Meng et al. [2018] in terms of model assumptions, since we incorporate review information from (former) employees for accurate salary prediction.

2.4 Topic Modeling for Opinion Analysis

Probabilistic topic models are capable of grouping semantic coherent words into human-interpretable topics. Archetypal topic models include probabilistic Latent Semantic Indexing (pLSI) [Hofmann 1999] and Latent Dirichlet Allocation (LDA) [Blei et al. 2003]. A lot of extensions have been proposed based on the above standard topic models, such as author-topic model [Rosen-Zvi et al. 2004], correlated topic model (CTM) [Blei and Lafferty 2005], dynamic topic model (DTM) [Blei and Lafferty 2006], and so on. Among them, numerous works focus on opinion analysis, especially for tackling the aspect-based opinion mining task [Vivekanandan and Aravindan 2014; Zhu et al. 2014]. Moreover, a few works have attempted to combine ratings and review texts when performing opinion analysis [Ganu et al. 2009; McAuley and Leskovec 2013; Titov and McDonald 2008].

As for the problem setting of opinion analysis, none of existing studies consider the simultaneous modeling of pros and cons texts during opinion profiling, which is one of our major concerns and is very important under the context of EBE.

2.5 Matrix Factorization for Prediction

Matrix factorization is a family of methods that is widely used for prediction. The intuition behind it is to get better data representation by projecting them into a latent space. Singular Value Decomposition (SVD) [Golub and Reinsch 1970] is a classic matrix factorization method for rating prediction in recommender systems, which gives low-rank approximations based on minimizing the sum-squared distance. However, SVD does not perform well on real-world sparse datasets with many missing values. To solve it, some probabilistic-based matrix factorization methods have been proposed [Marlin 2003; Marlin and Zemel 2004; Salakhutdinov and Mnih 2007; Zeng et al. 2015]. Among them, Probabilistic Matrix Factorization (PMF) [Salakhutdinov and Mnih 2007] is a representative one and is popular in industry. To improve prediction accuracy, Wang and Blei [2011] proposed to combine PMF with topic modeling of item content for scientific article recommendation, which is referred to as Collaborative Topic Regression (CTR). Based on CTR, many other approaches aim to incorporate more auxiliary information into the framework [Purushotham et al. 2012; Wang et al. 2013] or adapt deep learning techniques for content modeling [Li and She 2017; Wang et al. 2015; Zhang et al. 2016] and have shown their effectiveness in various real-world recommender systems. For instance, Wang et al. [2013] extended CTR by incorporating additional item social network information into the matrix factorization framework, while Purushotham et al. [2012] utilized users' social network information for recommendation. Wang et al. [2015] proposed a CDL model to jointly perform deep representation learning for content information and collaborative filtering for ratings matrix. Zhang et al. [2016] proposed a deep learning–based CKE model to leverage the heterogeneous item information in a knowledge base including structural content, textual content, and visual content for improved recommendation. Liu et al. [2017] developed a novel Bayesian inference algorithm making CTR capable of dealing with streaming data in an online learning setting. There also exist a stream of studies that focus on solving probabilistic matrix factorization in a non-linear way [Lawrence and Urtasun 2009]. Lawrence and Urtasun [2009] adopted a family of stochastic processes, namely, Gaussian Processes (GP) [Bonilla et al. 2010; Houlsby et al. 2012; Rasmussen and Williams 2005], which is designed to utilize changeable covariance functions and can therefore model complicated functions without limiting them to fixed form selected manually.

This line of method, especially CTR, inspires our unified approach to incorporate auxiliary review ratings, texts for more accurate salary prediction. However, our approach differs from CTR in several aspects. First, CTR is proposed for the purpose of content-based collaborative filtering, while our approach devises a novel perspective of treating review ratings and comments as complementary of each other, and side information for salary prediction. Second, our basic model CPCTR innovatively concerns with several key problems in EBE, e.g., simultaneous pros and cons modeling for opinion analysis, job-company pair modeling for more meaningful employer brand pattern mining. Third, our extended method GPCTR is among the earliest to utilize Gaussian processes for addressing the complex non-linear correlation issue across multiple heterogeneous information.

3 PRELIMINARIES

In this section, some preliminaries about research data, notations, and problem settings of EBE are presented.

3.1 Data Description

We collect the data from online employment platforms for EBE. Two page snapshots of Indeed,[1] a typical online recruitment platform, are illustrated in Figure 1 for better understanding. As can be seen from the review page for Microsoft (Figure 1(a)), for each employer (company), there are reviews published by its (former) staff, each containing the job position (e.g., Recreation Assistant) of the poster, textual comments about the pros (i.e., advantages), cons (i.e., disadvantages) of the target company, and a score rated from 1 to 5 indicating the poster's preference towards the company. Individual users are also encouraged to post their salary information to the platform anonymously. By collecting salary along with its corresponding job and company information from each individual, the platform can then calculate the average salary for each job position given a specific company, which finally forms the salary profile page as shown in Figure 1(b).

3.2 Notations

We assume that there are a total of E companies and J job positions. For company e, each review consists of the role of the corresponding reviewer (i.e., the reviewer's job position j), rating score $r_{j,e}$, and two independent textual segments, i.e., positive comments $\{w^P_{n,j,e}\}^N_{n=1}$ and negative comments $\{w^C_{m,j,e}\}^M_{m=1}$. Note that superscript character P denotes positive opinion and superscript character C denotes negative opinion. For each job-company pair (j, e), the average salary is denoted as $s_{j,e}$. We put the detailed mathematical notations used throughout the article in Table 1.

3.3 Problem Settings

Given the reviews and salaries from online recruitment platforms, two major goals are concerned to perform holistic EBE in this article: (1) how to learn good patterns from opinion texts for profiling of targeted jobs and companies; (2) how to conduct effective salary benchmarking for predicting missing salary information in the job market. Salary benchmarking (prediction) is the focal point for EBE, since both employers and applicants can benefit substantially from quantifying the

[1] http://www.indeed.com/.

Table 1. Mathematical Notations

Symbol	Description
K	The predefined number of topics.
J	The number of unique job positions in the recruitment markets.
E	The number of companies in the markets.
N	The number of words in positive review document for job-company pair (j, e).
M	The number of words in negative review document for job-company pair (j, e).
$r_{j,e}$	The average observed rating value for job-company pair (j, e).
$s_{j,e}$	The average observed salary value for job-company pair (j, e).
	CPCTR
$\beta_{k,j}$	Positive topic-word distribution of topic k for job position j.
$\varphi_{k,j}$	Negative topic-word distribution of topic k for job position j.
$\theta_{j,e}$	Topic mixture proportion for job-company pair (j, e).
$w^P_{n,j,e}$	Term indicator of the nth word in positive review document for job-company pair (j, e).
$w^C_{m,j,e}$	Term indicator of the mth word in negative review document for job-company pair (j, e).
$z^P_{n,j,e}$	Latent topic assignment of the nth word in positive review document for job-company pair (j, e).
$z^C_{m,j,e}$	Latent topic assignment of the mth word in positive review document for job-company pair (j, e).
b_j	Latent factor vector that governs generation of ratings for job position j.
u_j	Latent factor vector that governs generation of salaries for job position j.
$v_{j,e}$	Latent factor vector for job-company pair (j, e).
$\epsilon_{j,e}$	Latent offset vector that bridges $v_{j,e}$ and $\theta_{j,e}$.
α	Hyperparameter of the Dirichlet prior on $\theta_{j,e}$.
λ_v	Hyperparameter of the Gaussian prior on $\epsilon_{j,e}$.
λ_s	Hyperparameter of the Gaussian distribution that draws $s_{j,e}$.
λ_r	Hyperparameter of the Gaussian distribution that draws $r_{j,e}$.
λ_u	Hyperparameter of the Gaussian prior on u_j.
λ_b	Hyperparameter of the Gaussian prior on b_j.
	GPCTR
β_k	Positive topic-word distribution of topic k.
φ_k	Negative topic-word distribution of topic k.
θ_j	Topic mixture proportion for job j.
$w^P_{n,j}$	Term indicator of the nth word in positive review document for job j.
$w^C_{m,j}$	Term indicator of the mth word in negative review document for job j.
$z^P_{n,j}$	Latent topic assignment of the nth word in positive review document for job j.
$z^C_{m,j}$	Latent topic assignment of the mth word in positive review document for job j.
v_j	Latent factor vector for job j.
ϵ_j	Latent offset vector that bridges v_j and θ_j.
α	Hyperparameter of the Dirichlet prior on θ_j.
λ_v	Hyperparameter of the Gaussian prior on ϵ_j.
α_s, λ_s	Parameters of covariance function of Gaussian processes for generating salary.
α_r, λ_r	Parameters of covariance function of Gaussian processes for generating rating.

salary expectation with respect to different jobs and companies. We tackle salary benchmarking in a matrix completion manner; that is, to predict missing entry $\hat{s}_{j,e}$ in a constructed job-company salary matrix $\{s_{j,e}\}_{j=1,e=1}^{J,E}$. We believe that opinion profiling and salary benchmarking in the context of EBE are of great interest to both academia and industry.

4 PROPOSED APPROACH

In this section, we first introduce the general ideas of the proposed approach and then give the technical details of our models.

4.1 General Ideas

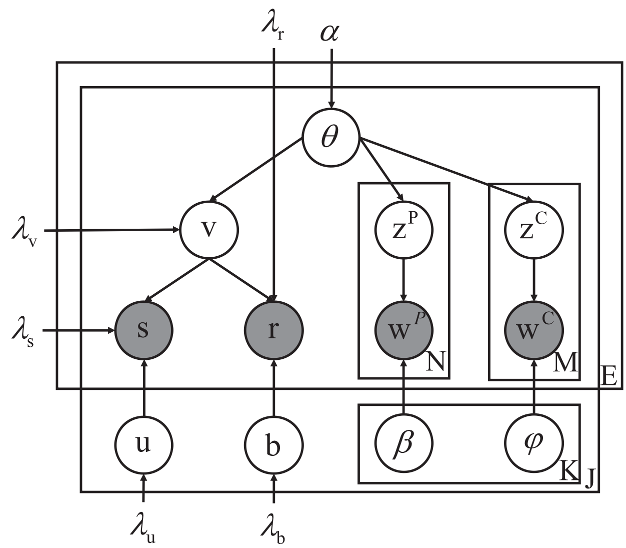
To achieve the goals of EBE, two key machine learning challenges need to be addressed as follows: First, the constructed job-salary matrix is highly sparse. This poses great difficulty in the task of matrix completion. An intuitive solution is to incorporate more information for accurate salary prediction. In our settings, we treat review ratings and comments as complementary of each other, both of which are in turn regarded as side information for salary prediction. To this end, we propose a basic method CPCTR for jointly modeling textual comments, numerical ratings, and salaries. Specifically, we employ topic models for learning opinion patterns from texts and conduct probabilistic matrix factorization over numerical values. By this means, CPCTR is capable of performing opinion profiling and salary prediction simultaneously. Second, the heterogeneity among different information, especially the non-linear correlation, indeed hinders the power of fusion method. To better capture the non-linearity across different data sources, we further propose an extended Gaussian processes–based model GPCTR for enhancing salary prediction performance. We first describe the technical details of the basic model CPCTR in Section 4.2 and then introduce the extended model GPCTR in Section 4.3.

4.2 Company Profiling–based Collaborative Topic Regression

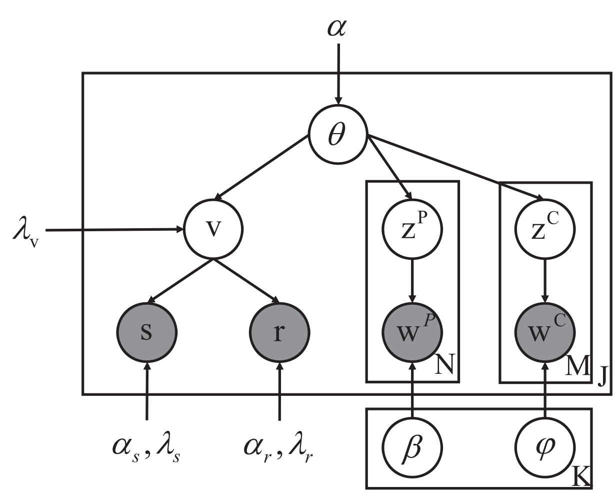
As mentioned above, our proposed CPCTR model is a joint model that combines topic modeling with matrix factorization. The graphical representation of CPCTR is shown in Figure 2(a), and the complete generative process of our model is demonstrated in Algorithm 1. To facilitate understanding, we look into the model on two sides.

On the one hand, we model the job-company pair with a latent topic vector $\theta_{j,e} \in R^K$, where K is the number of topics. In probabilistic topic modeling, job position j can be represented by two latent matrices, i.e., the positive opinion topics $\beta_j \in R^{K\times G}$ and the negative opinion topics $\varphi_j \in R^{K\times G}$, where G is the size of vocabulary. For the nth word $w^P_{n,j,e}$ in a positive review of job-company pair, we assume there is a latent variable denoted as $z^P_{n,j,e} \in \{1,\ldots,K\}$, indicating the word's corresponding topic. To be more specific, given $z^P_{n,j,e} = k$, $w^P_{n,j,e}$ follows a multinomial distribution parameterized by $\beta_{k,j}$. Meanwhile, the positive latent pattern $z^P_{n,j,e}$ is considered to be drawn from the multinomial distribution $Mult(\theta_{j,e})$. A similar process can be conducted for the negative review.

On the other hand, we conduct matrix factorization for generating ratings and salaries. In matrix factorization, we represent job position and job-company pair in a shared latent low-dimensional space of dimension K, i.e., job position j is represented by latent vectors $u_j \in R^K$ and $b_j \in R^K$, which indicate the influences of job positions over salaries and ratings, respectively. Similarly, the job-company pair (j,e) is represented by a latent vector $v_{j,e} \in R^K$, which indicates the joint influences of job-company pair over numeric ratings and salaries. Here, we assume the latent vector u_j and b_j follow Gaussian distributions with parameters $(0, \lambda_u^{-1}\mathbf{I}_K)$ and $(0, \lambda_b^{-1}\mathbf{I}_K)$, respectively. And, the latent vector $v_{j,e}$ is derived from $\theta_{j,e}$ by adding an offset ϵ, which also follows a Gaussian



(a) CPCTR.



(b) GPCTR.

Fig. 2. The graphical representation of our proposed models.

distribution with parameters $(0, \lambda_v^{-1}\mathbf{I}_K)$. Therefore, it is obvious that $v_{j,e}$ is the key point by which we jointly model both content and numerical information.

We form salary prediction of a specific job-company pair $(\hat{s}_{j,e})$ through the inner product between their latent representations, i.e.,

$$\hat{s}_{j,e} = u_j^\top v_{j,e}. \tag{1}$$

For parameter learning, we follow a similar procedure to CTR [Wang and Blei 2011]. We leverage a variational Expectation-Maximization (EM) method for parameter learning of CPCTR.

ALGORITHM 1: The Generative Process of CPCTR

(1) For each job position j,
  (a) draw latent vector $u_j \sim N(0, \lambda_u^{-1}\mathbf{I}_K)$.
  (b) draw latent vector $b_j \sim N(0, \lambda_b^{-1}\mathbf{I}_K)$.
(2) For each job-company pair (j, e),
  (a) Draw topic proportion $\theta_{j,e}$ from the Dirichlet prior $Dir(\alpha)$.
  (b) For the nth word $w_{n,j,e}^P$ of positive review,
    (i) Draw topic assignment $z_{n,j,e}^P \sim Mult(\theta_{j,e})$.
    (ii) Draw word $w_{n,j,e}^P \sim Mult(\beta_{z_{n,j,e}^P, j})$.
  (c) For the mth word $w_{m,j,e}^C$ of negative review,
    (i) Draw topic assignment $z_{m,j,e}^C \sim Mult(\theta_{j,e})$.
    (ii) Draw word $w_{m,j,e}^C \sim Mult(\varphi_{z_{m,j,e}^C, j})$.
  (d) Draw latent offset $\epsilon_{j,e} \sim \mathcal{N}(0, \lambda_v^{-1}\mathbf{I}_K)$ and set latent vector $v_{j,e} = \epsilon_{j,e} + \theta_{j,e}$.
  (e) Draw rating/salary values,
    (i) draw rating value $r_{j,e} \sim \mathcal{N}(b_j^\top v_{j,e}, \lambda_r^{-1}\mathbf{I}_K)$.
    (ii) draw salary value $s_{j,e} \sim \mathcal{N}(u_j^\top v_{j,e}, \lambda_s^{-1}\mathbf{I}_K)$.

In the E-step, we can update the latent vector of u_j, b_j and $v_{j,e}$ as follows:

$$u_j = \left(\lambda_u \mathbf{I}_K + \lambda_s \sum_e v_{j,e} v_{j,e}^\top\right)^{-1} \lambda_s \sum_e v_{j,e} s_{j,e}, \tag{2}$$

$$b_j = \left(\lambda_b \mathbf{I}_K + \lambda_r \sum_e v_{j,e} v_{j,e}^\top\right)^{-1} \lambda_r \sum_e v_{j,e} r_{j,e}, \tag{3}$$

$$v_{j,e} = (\lambda_v \mathbf{I}_K + \lambda_s u_j u_j^\top + \lambda_r b_j b_j^\top)^{-1} (\lambda_v \theta_{j,e} + \lambda_s u_j s_{j,e} + \lambda_r b_j r_{j,e}). \tag{4}$$

For learning $\theta_{j,e}$, we introduce variational distributions $q(z_{n,j,e}^P = k) = \phi_{k,n,j,e}^P$ and $q(z_{m,j,e}^C = k) = \phi_{k,m,j,e}^C$. We update $\phi_{k,n,j,e}^P$ and $\phi_{k,m,j,e}^C$ as follows:

$$\phi_{k,n,j,e}^P \propto \theta_{k,j,e} \beta_{k, w_{n,j,e}^P}, \tag{5}$$

$$\phi_{k,m,j,e}^C \propto \theta_{k,j,e} \varphi_{k, w_{m,j,e}^C}. \tag{6}$$

With $\phi_{k,n,j,e}^P$ and $\phi_{k,m,j,e}^C$ updated, we can use projection gradient [Bertsekas 1999] to optimize $\theta_{j,e}$.

In the M-step, we optimize β and φ as follows:

$$\beta_{g,k,j} \propto \sum_e \sum_n \phi_{k,n,j,e}^P \mathbb{I}\left[w_{n,j,e}^P = g\right], \tag{7}$$

$$\varphi_{g,k,j} \propto \sum_e \sum_m \phi_{k,m,j,e}^C \mathbb{I}\left[w_{m,j,e}^C = g\right], \tag{8}$$

where we denote g as an arbitrary term in the vocabulary set. The detailed derivations of CPCTR's parameter learning procedure can be found in Appendix A.1.

Comparison to CTR. First, CPCTR treats review ratings and comments as complementary of each other, both of which are in turn regarded as side information for salary benchmarking (prediction). This is a novel perspective compared to CTR in the context of EBE and is indeed crucial to accurate salary benchmarking. Second, unlike CTR, which only learns a global topic-word distribution, CPCTR can learn two kinds of job-related topic-word patterns, including a positive topic-word distribution β_j and a negative topic-word distribution φ_j. This design enables analysis of pros and cons, which are very important for EBE. Finally, different from CTR assuming each job has a topic proportion, CPCTR assumes that each job-company pair has a topic proportion. This makes CPCTR capable of producing fine-grained semantics (namely, most probable words for each job-company pair) and learning more meaningful patterns for accurately predicting salary.

4.3 Gaussian Processes–based Collaborative Topic Regression

By the joint modeling of review texts, ratings, and salaries with CPCTR model, we have achieved to establish the intrinsic connection of three data sources from a probabilistic perspective. However, in real-world practice, the great challenge lies in that different data sources are usually very noisy, and the correlation inside/across different data sources might be complex and arbitrary. In light of this, here, we provide an extension model of CPCTR, denoted as GPCTR (Gaussian Processes–based Collaborative Topic Regression), which introduces stochastic Gaussian processes to capture the complex correlation hidden inside latent spaces. The graphical representation of GPCTR is shown in Figure 2(b).

It is indeed very important in our framework to bridge the gap between textual information with numeric information (e.g., CPCTR bridges θ and v with latent offset vectors ϵ), which might need subtle modeling of non-linear correlation patterns among this heterogenous information. Fortunately, Gaussian processes are designed to utilize changeable covariance functions and can therefore model complicated functions without limiting them to fixed form selected manually. This raises the opportunity for tackling the above issue.

A Gaussian process (GP) is a nonparametric model in Bayesian statistics [Rasmussen and Williams 2005]. In general, a GP is usually used as priors, which can be seen as a probability distribution for generating functions. Formally, we assume a random real-valued function $f : \mathbb{R}^K \rightarrow \mathbb{R}$, mapping K-dimensional real-valued feature vector into real-valued scalar. The function f follows a GP, denoted as $f \sim \mathcal{GP}(H, C)$, if $\boldsymbol{f}$ meets the following conditions: For any finite-size feature matrix $X \in \mathbb{R}^{N \times K}$, the joint distribution of $f = \{f(x_n)\}_{n=1}^{N}$ over those observations X follows a multivariate Gaussian distribution, i.e., $p(f|X) = \mathcal{N}(H, C)$. Note that H and C are the mean and covariance function for this multivariate Gaussian distribution. Particularly, the covariance function measures the degree of pairwise correlation between two samples in the feature matrix X. Practically, we fix the mean function H with a zero mean function, i.e., $\mathbf{H} \equiv \mathbf{0}$, and assume the covariance function C to take a parametric form, e.g., the linear covariance function takes the form $c(x_{i,:}, x_{j,:}) = x_{i,:}^{\top} x_{j,:}$ while the non-linear RBF covariance function takes the form $c(x_{i,:}, x_{j,:}) = \alpha \exp(-\frac{\gamma}{2}||x_{i,:} - x_{j,:}||^2)$.

As such, instead of using conventional matrix factorization technique, an intuitive way to better gap different data views is to alternatively use Gaussian processes to map latent feature vectors (i.e., v) to the final predicted numeric information (i.e., ratings and salaries). To this end, we resort to GP latent variable models (NLMFGP) [Lawrence and Urtasun 2009; Rasmussen and Williams 2005] that are particularly effective in capturing arbitrary correlation (e.g., linear or non-linear) hidden inside latent feature space.

Before going into more details of the extension model GPCTR, we revisit another issue with CPCTR: The proposed CPCTR model suffers from high computational cost. As mentioned in above sections, CPCTR jointly models review texts, ratings, and salaries. It associates each job-company pair (j, e) with a topic proportion vector $\theta_{j,e}$, a latent vector $v_{j,e}$, and assumes that each job position j has two topic-word distributions β_j, φ_j. This indicates that in CPCTR the number of latent vectors θ, v scale linearly with $E \times J$ and the number of latent vectors β, φ scale linearly with J. Moreover, the model complexity and the inference difficulty improve considerably if the Gaussian processes are further introduced to our extension model design. To address this issue, we first extend the proposed CPCTR model into a simplified form. Specifically, in the extension model, each job position j is associated with a topic proportion θ_j and the positive/negative topic-word distributions are governed by two global K-dimensional parameters β and φ.

Therefore, in terms the generative process of review texts, we first introduce a simplified form of the original framework:

For each job position j,

(a) Draw topic proportion θ_j from the Dirichlet prior $Dir(\alpha)$.
(b) For the nth word $w^P_{n,j}$ of positive review,
  (i) Draw topic assignment $z^P_{n,j} \sim Mult(\theta_j)$.
  (ii) Draw word $w^P_{n,j} \sim Mult(\beta_{z^P_{n,j}})$.
(c) For the mth word $w^C_{m,j}$ of negative review,
  (i) Draw topic assignment $z^C_{m,j} \sim Mult(\theta_j)$.
  (ii) Draw word $w^C_{m,j} \sim Mult(\varphi_{z^C_{m,j}})$.

For the generation of numeric ratings and salaries, we start the derivation within the context of non-linear PMF [Lawrence and Urtasun 2009]. Rewriting the probabilistic form of conventional PMF for generating salaries, we can obtain:

$$p(S|V, U, \lambda_s) = \prod_{j=1}^{J} \prod_{e=1}^{E} \mathcal{N}\left(s_{j,e} | u_e^\top v_j, \lambda_s^{-1} \mathbf{I}_K\right), \tag{9}$$

$$p(V|\theta, \lambda_v) = \prod_{j=1}^{J} \prod_{k=1}^{K} \mathcal{N}\left(v_{j,k} | \theta_{j,k}, \lambda_v^{-1}\right), \tag{10}$$

where $V \in \mathbb{R}^{J \times K}, U \in \mathbb{R}^{E \times K}$, denote the latent job and company matrix, respectively. Here, similar to CPCTR, we place a Gaussian offset to the priors of v to bridge the gap between topic proportion θ and v. We omit the priors over U, since we can further use $f_e(v_j) = u_e^\top v_j$ to rewrite Equation (9) as follows:

$$p(S|V, f, \lambda_s) = \prod_{j=1}^{J} \prod_{e=1}^{E} \mathcal{N}(s_{j,e} | f_e(v_j), \lambda_s^{-1} \mathbf{I}_K). \tag{11}$$

By placing priors over function $f(\cdot)$ with a GP, we can recover the NLMFGP as follows:

$$p(f|V) = \mathcal{N}(\mathbf{0}, C), \tag{12}$$

and marginalizing f leads to

$$p(S|V, \alpha_s, \lambda_s) = \prod_{e=1}^{E} \mathcal{N}(s_e | \mathbf{0}, \alpha_s^{-1} V V^\top + \lambda_s^{-1} \mathbf{I}_J), \tag{13}$$

where the covariance function is linear, and α_s, λ_s are the parameters for the covariance function.

In real-world applications, the data sparsity problem arises, e.g., in the salary prediction task, we can only observe partial salaries in the matrix S. Hereinafter, we denote the observed subset of S as S_o, where o is the observed values' index, $s_{o_e,e}$ represents the observed set for company e, and V_{o_e} is its corresponding latent job matrix. The likelihood would have the form

$$p(S|V, \alpha_s, \lambda_s) = \prod_{e=1}^{E} \mathcal{N}\left(s_{o_e,e}|\mathbf{0}, \alpha_s^{-1} V_{o_e} V_{o_e}^\top + \lambda_s^{-1}\mathbf{I}\right). \tag{14}$$

Moreover, we give the derivation of generating rating matrix R in a similar manner:

$$p(R|V, \alpha_r, \lambda_r) = \prod_{e=1}^{E} \mathcal{N}\left(r_{o_e,e}|\mathbf{0}, \alpha_r^{-1} V_{o_e} V_{o_e}^\top + \lambda_r^{-1}\mathbf{I}\right), \tag{15}$$

where α_r, λ_r are the corresponding parameters for the covariance function for generating rating values.

For predicting the salaries of a new job-company pair (j, e), we apply the standard formula for Gaussian processes prediction:

$$\hat{s}_{j,e} = \left[\left(\alpha_s^{-1} V_{o_e} V_{o_e}^\top + \lambda_s^{-1}\mathbf{I}\right)^{-1} c_{o_e,j}\right]^\top S_{o_e}. \tag{16}$$

For parameter learning, we train our models on the observed set S_o. The parameter learning procedure takes a similar manner as introduced for CPCTR, except for the udpate of the latent matrix V, which is estimated with Stochastic Gradient Descent (SGD). Details about the parameter learning procedure of GPCTR are described in Appendix A.2.

5 EXPERIMENTAL RESULTS

In this section, we first qualitatively evaluate the positive and negative opinion patterns learned from review texts posted by employees on Kanzhun, then quantitatively validate the effectiveness and robustness of the proposed CPCTR model and its extension GPCTR in the task of salary prediction. Additional experiments are conducted to verify the key components in CPCTR. Ablation experiments are designed to study different data sources' contribution to the predictive performance of GPCTR. Moreover, we show the capabilities of our models in capturing correlations between numeric ratings and salaries. Besides, we give sensitivity analysis of the proposed models to some important hyperparameters. Finally, to show the generalizability of our proposed models over areas other than EBE, we apply our methods to Amazon product brand evaluation and empirically evaluate the performance of our models on the aspect of product price prediction.

5.1 Experimental Setup

In this section, we first introduce the basic data statistics of the two datasets, i.e., Kanzhun and Glassdoor, which are later used to evaluate the performance for salary prediction. Then, we describe the baseline methods and evaluation protocol for comparison.

5.1.1 Datasets and Statistics. We empirically evaluate the salary prediction performance on two real-world datasets, including Kanzhun and Glassdoor. Similar to Indeed as shown in Figure 1, Kanzhun[2] and Glassdoor[3] are two of the largest online recruitment websites in China and the U.S., respectively, where individual users can rate job positions of a specific company as its (former) employees and post review texts (i.e., pros and cons) on the corresponding job positions. Individuals are also encouraged to post their job salary information to these platforms, which can later be aggregated by the platforms to create a partial salary profile of specific company-job pairs (e.g.,

[2] http://www.kanzhun.com/.
[3] https://www.glassdoor.com/index.htm.

Table 2. Statistics of Datasets for Employer Brand Evaluation

Statistics	Kanzhun	Glassdoor
# of Companies	2,503	1,007
# of Jobs	2,013	4,619
# of Salaries/Ratings/Reviews	16,641	12,352
# of Pros Words	466,661	761,140
# of Cons Words	250,446	952,828
Matrix Sparsity	0.9967	0.9973
Average Non-zero Entries per Company	6.65	12.27
Average Pros Words per Job	231.82	164.78
Average Cons Words per Job	124.41	206.28

average salaries or salary ranges). Thus, both the two platforms provide ideal data sources for experiments on employer competitiveness analysis and salary benchmarking, which are in accordance with our problem settings. We crawled salary and review pages from the website of Kanzhun and Glassdoor and then extracted review ratings, texts, and salaries from the raw pages, forming the two datasets. The ratings on both datasets range from 1 to 5. The main difference of the two datasets lies in that the review texts on Kanzhun are mainly Chinese characters while those on Glassdoor are in English.

We show the statistics of the two datasets in Table 2. Concerning review texts, we group reviews by their job positions and thus obtain one aggregated document for each job-company pair. Then, we remove single words/stop words and select top 10K frequent words as vocabulary set. Finally, we convert documents into the bag-of-words format for model learning. For numeric ratings/salaries, we calculate the average values for each job-company pair and thus construct numeric matrices like in collaborative filtering tasks. For simplicity, we align different data sources and reserve the job-company pairs, which have non-zero entries of ratings and salaries. It can be seen that rating and salary matrices are really sparse, and the average document length is rather short on both datasets. For example, the average number of cons words for each job on Kanzhun is 124.41. For ease of model training, min-max method is adopted to normalize all ratings/salaries into [0, 1] range.

5.1.2 Baseline Methods. For evaluating our proposed methods' performance in salary prediction, we conduct experiments to predict unobserved salaries given some partially observed salaries, which is similar to conventional Collaborative Filtering (CF) tasks. For comparison, we chose two categories of methods as baselines: CF methods and regression-based methods.

The three CF methods are as follows:

- **PMF (Probabilistic Matrix Factorization)** [Salakhutdinov and Mnih 2007]: This method adopts matrix factorization technique for CF from a probabilistic perspective. As a state-of-the-art method for latent factors CF, PMF performs competitively on extremely sparse datasets. This method conducts matrix factorization only on salary information. We implement this method in Python with reference to the Matlab code[4] provided by Ruslan Salakhutdinov.

[4]http://www.cs.toronto.edu/~rsalakhu/BPMF.html.

- **NLMFGP (Non-linear Matrix Factorization with Gaussian Processes)** [Lawrence and Urtasun 2009]: This method uses nonlinear matrix factorization via GP. It conducts matrix factorization only on salary information. We use the source code implemented by Python[5].
- **CTR (Collaborative Topic Regression)** [Wang and Blei 2011]: This method combines CF and probabilistic topic models. CTR uses both salary information and review texts. We implement this method in Python with reference to source code[6] provided by Chong Wang.

The two regression-based methods are as follows:

- **LR (Linear Regression)** [Galton 1886]: This method fits a linear regressor model, which learns a function mapping review text and rating of each job-company pair to its corresponding salary. We use the Python implementation included in Scikit-learn library.[7]
- **DTR (Decision Tree Regressor)** [Breiman et al. 1984]: This method fits a decision tree regressor model, which learns simple decision rules inferred from input features (i.e., review text and rating) and output (i.e., salary). We also use the Python implementation included in Scikit-learn library.[8]

5.1.3 Evaluation Protocol. In our experiments, we adopted a company-oriented generalization technique for evaluating the performance of different methods on held-out dataset, which can be seen as weak generalization, as described in Marlin's setup [Marlin 2004] for CF tasks. Specifically, for each company in the salary matrix, we randomly split $p\%$ entries for training and the predictive results are reported on the remaining $1 - p\%$ during the testing phase. To validate the robustness of different models, we vary p in $\{10, 20, 30, 40, 50, 60, 70, 80, 90\}\%$. For each setting of p, we conduct five times independent random train-test split and then the reported results are averaged over five times independent run.

5.1.4 Evaluation Metric. For measuring the predictive performance of different methods in salary prediction, we chose a popular metric, i.e., Root-Mean-Square Error (RMSE). This metric is calculated as follows:

$$\mathrm{RMSE}_{\mathcal{T}} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(j,e) \in \mathcal{T}} (s_{j,e} - \hat{s}_{j,e})^2}, \tag{17}$$

where $\mathcal{T}$ is the set for testing data samples, $s_{j,e}$ is the actual value of salary for job-company pair (j, e), and $\hat{s}_{j,e}$ is its corresponding predicted value.

5.1.5 Parameter Settings for Comparison. For all methods, grid search is used to find best parameters. Specifically, when doing grid search, 5-fold cross-validation on the training set with $p = 80\%$ is conducted to evaluate the model's performance. Note that we do the parameter selection process only on Kanzhun dataset and fix the best parameters as the default parameter settings for all experiments.

The default parameter settings of different methods are stated as follows: For all methods, we set the number of latent factors to $K = 5$ and the maximum iterations for convergence as max_iter = 50. For CTR, we set $\lambda_u = 0.1$, $\lambda_v = 100.0$, $a = 1.0$, and $b = 0.01$. For PMF, we find that momentum = 0.9, learning rate = 0.001 gives good performance, which is consistent with the empirical findings in Salakhutdinov and Mnih [2007]; training data are randomly shuffled prior to each epoch. For CPCTR, we set $\lambda_v = 10.0$, $\lambda_r = 1.0$, $\lambda_s = 1.0$, $\lambda_u = 0.1$, $\lambda_b = 0.01$. According

[5]https://github.com/Salma-El-Alaoui/Recommender-Systems.
[6]https://github.com/blei-lab/ctr.
[7]https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html.
[8]https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html.

Table 3. Topics of Job Position "Software Engineer"

Positive Topics			Negative Topics		
Topic 0	Topic 1	Topic 2	Topic 0	Topic 1	Topic 2
环境(environment)	轻松(relaxed)	技术(technology)	加班(overtime)	前景(prospect)	技术(technology)
不错(good)	福利(welfare)	待遇(treatment)	流程(flow)	环境(environment)	机会(opportunity)
机会(opportunity)	工资(salary)	机会(opportunity)	需要(need)	提升(improve)	太低(low)
成长(growth)	弹性(flexible)	环境(environment)	发展(improve)	氛围(atmosphere)	发展(improve)
氛围(atmosphere)	相处(get along)	部门(department)	时间(time)	待遇(treatment)	福利(welfare)

Table 4. Pros and Cons of Various Enterprises Given Job Position "Software Engineer"

百度(Baidu)		阿里巴巴(Alibaba)		腾讯(Tencent)	
优势(Pros)	劣势(Cons)	优势(Pros)	劣势(Cons)	优势(Pros)	劣势(Cons)
福利(welfare)	需要(need)	福利(welfare)	压力(pressure)	锻炼(training)	加班(overtime)
氛围(atmosphere)	加班(overtime)	氛围(atmosphere)	重复(repeat)	学到(learning)	压力(pressure)
轻松(relaxed)	严重(serious)	文化(culture)	加班(overtime)	不错(good)	管理(management)
技术(technology)	压力(pressure)	待遇(treatment)	劳动(labor)	空间(space)	公司(company)
环境(environment)	管理(management)	应届生(freshgraduates)	销售(sales)	和谐(harmonious)	前景(prospect)

to Lawrence and Urtasun [2009], since linear kernel achieves competitive performance as non-linear (e.g., RBF) kernel—however, with much less training complexity—we apply linear kernel for GPCTR. Moreover, to train GPCTR model, we used Stochastic Gradient Descent (SGD) and set the learning rate to 1×10^{-4} with a momentum setting to 0.9. According to empirical rules, we initialized all covariance parameters as 1 except noise and bias variance. Noise variance was set to 5 and bias variance was set to 0.11. λ_v was set to 0.001. For initializing latent factors V, we drew from a spherical Gaussian distribution with zero mean and standard deviation 1×10^{-3}. Discussions about parameter sensitivity of CPCTR and GPCTR are given in Section 5.7.

5.2 Empirical Study of Opinion Profiling

Here, we apply CPCTR to carry out opinion analysis for different companies based on employees' reviews on Kanzhun. The objective is to effectively reveal the pros and cons of companies, which indeed helps for employer competitiveness evaluation.

To illustrate the effectiveness of learning job-position-level topic-word distributions, we listed three positive topics and three negative topics of job position *Software Engineer* inferred from CPCTR, as shown in Table 3. Each topic is represented by five most probable words for that topic. It can be seen that our method has an effective interpretation of latent job position pattern and these topics accurately capture the common semantics of the job position *Software Engineer* in the whole market. We can see some interesting positive/negative topic patterns. For positive topics, topic 0 is about job environment, topic 1 is about flexible work time, topic 2 is technology atmosphere. For negative topics, topic 0 is about overtime, topic 1 is about prospect and promotion, and topic 2 is about opportunity and welfare.

We also compared the pros and cons among BAT, which is the abbreviation of three largest and most representative Chinese Internet companies, i.e., Baidu, Alibaba, and Tencent. Specifically, we presented the pros and cons with most probable words appearing in learned topics for each company, given the job position *Software Engineer* in Table 4. As can be seen, topics for each job-company pair can effectively capture the specific characteristics of each company. For instance,

Table 5. Pros and Cons of Various Job Positions for Baidu

用户体验设计师(UE Designer)		办公室文员(Office Administrator)		销售经理(Sales Manager)	
优势(Pros)	劣势(Cons)	优势(Pros)	劣势(Cons)	优势(Pros)	劣势(Cons)
平台(platform)	待遇(treatment)	学习(learning)	职责(duty)	年假(vacation)	压力(pressure)
朋友(friendship)	效率(efficiency)	应届生(freshgraduates)	房租(rent)	福利(welfare)	晋升(promotion)
学到(learning)	加班(overtime)	专业(professional)	消费(consumption)	事业(fulfillment)	期权(option)
晋升(promotion)	痛苦(suffering)	毕业生(graduate)	繁重(heavy)	调休(compensation)	流程(procedure)
发展(development)	休息(rest)	平台(platform)	迷茫(depressed)	氛围(atmosphere)	流动(flow)

the typical pros of Baidu are welfare and technology, while those of Tencent are training and learning and those of Alibaba are culture and atmosphere. Interestingly, employees of all these three companies chose overtime as their cons, and the management of Tencent seems to be a typical cons.

Apart from comparison among different companies, we additionally took Baidu as an example and conducted competitiveness analysis over its different job positions, including *User Experience Designer*, *Office Administrator,* and *Sales Manager*. We chose these three job positions because they may represent three typical roles in an IT company. The pros and cons of the three job positions for Baidu are shown in Table 5. From the results, we can see that different roles of employees in the same company may have different opinion focus towards their employer. For example, *Sales Manager* may prefer to pay more attention to the aspect closely related to enterprise welfare (e.g., vacation, compensation, and option) while *Office Administrator* may have more focus on the aspect of living cost (e.g., room rent and living consumption).

5.3 Performance of Salary Prediction

We evalute the predictive performance of different models on the two datasets in Table 6 and Table 7, respectively. The average value of RMSE and its standard deviation in five times independent run are shown. We highlight the best results in bold and mark the runner-up with underline.

For clarity of comparison, we also include the performance improvement of GPCTR over other methods in terms of RMSE and calculate the average improvement with varying size of training data in the last column. Moreover, we perform Turkey's Honestly Significant Difference (HSD) test [Sakai 2018].

From the results, we have the following observations:

- First, with increasing percentage of the training set from 10 to 90 on the two datasets, most CF-based methods achieve improved prediction performance, i.e., decreasing value of RMSE. This indicates that CF-based methods benefit from more training data.
- Second, by integrating three data sources, including review opinions (i.e., texts), ratings, and salaries, our methods considerably improve salary prediction performance in terms of RMSE on both two datasets. It can be seen that the proposed integrated models (i.e., CPCTR and GPCTR) outperform the other three competitive baselines in most of the experimental settings, especially on Kanzhun dataset. For example, GPCTR surprisingly outperforms PMF by an average 54.48% relative RMSE improvement on Glassdoor. This greatly demonstrates the potential power of the framework of integrating review texts and numerical ratings for accurate salary prediction. The results of Turkey's HSD test also shows the outperformance in most cases is statistically significant with p-value < 0.05.
- It can also be seen that the extension model GPCTR always outperforms the proposed CPCTR model on the two datasets. On Kanzhun, GPCTR outperforms CPCTR by an

Table 6. Performance of Salary Prediction on Kanzhun Dataset

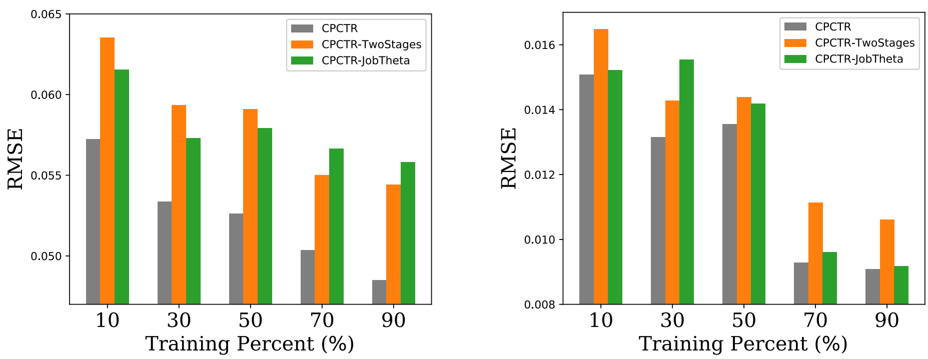
Training (%)		10	20	30	40	50	60	70	80	90	Average
Methods	LR	0.06047 ±0.0020	0.06272 ±0.0009	0.06347 ±0.0011	0.06685 ±0.0011	0.07133 ±0.0015	0.07525 ±0.0032	0.08671 ±0.0023	0.09189 ±0.0026	0.10912 ±0.0041	—
	DTR	0.06065 ±0.0018	0.06103 ±0.0022	0.06007 ±0.0007	0.06081 ±0.0012	0.05994 ±0.0014	0.05868 ±0.0009	0.05792 ±0.0020	0.05747 ±0.0021	0.05911 ±0.0044	—
	CTR	0.07487 ±0.0005	0.07309 ±0.0008	0.06999 ±0.0009	0.06976 ±0.0008	0.06803 ±0.0009	0.06598 ±0.0012	0.06408 ±0.0014	0.06368 ±0.0020	0.06212 ±0.0033	—
	PMF	**0.05513 ±0.0004**	0.05498 ±0.0003	0.05460 ±0.0006	0.05564 ±0.0004	0.05611 ±0.0006	0.05456 ±0.0011	0.05474 ±0.0015	0.05420 ±0.0015	0.05386 ±0.0030	—
	NLMFGP	0.06329 ±0.0018	0.05964 ±0.0019	0.06719 ±0.0071	0.07224 ±0.0034	0.06391 ±0.0081	0.07524 ±0.0186	0.05148 ±0.0010	0.06547 ±0.0182	0.04934 ±0.0032	—
	CPCTR	*0.05725 ±0.0003*	0.05514 ±0.0004	*0.05337 ±0.0007*	*0.05336 ±0.0007*	*0.05261 ±0.0004*	*0.05033 ±0.0012*	*0.05036 ±0.0019*	*0.04961 ±0.0024*	*0.04850 ±0.0036*	—
	GPCTR	0.06001 ±0.0004	**0.05479 ±0.0005**	**0.05090 ±0.0006**	**0.05051 ±0.0006**	**0.04970 ±0.0008**	**0.04775 ±0.0007**	**0.04703 ±0.0015**	**0.04589 ±0.0023**	**0.04479 ±0.0033**	—
Improvement of GPCTR vs.	LR(%)	0.76	12.64*	19.80*	24.44*	30.32*	36.54*	45.76*	50.06*	58.95*	31.03
	DTR(%)	1.06	10.22*	15.27*	16.93*	17.08*	18.63	18.80*	20.15	24.23*	15.82
	CTR(%)	19.85*	25.04*	27.28*	27.59*	26.94*	27.63*	26.61*	27.94*	27.90*	26.31
	PMF(%)	-8.85	0.35	6.78	9.22*	11.42*	12.48	14.08*	15.33	16.84*	8.63
	NLMFGP(%)	5.18*	8.13*	24.24*	30.08*	22.23*	36.54	8.64*	29.90	9.22	19.35
	CPCTR(%)	-4.82	0.63	4.63	5.34*	5.53*	5.13*	6.61*	7.50	7.65	4.24

Note: The significant improvements of GPCTR over the comparative methods are marked by star superscript (Turkey's HSD, $p < .05$).

Table 7. Performance of Salary Prediction on Glassdoor Dataset

Training (%)		10	20	30	40	50	60	70	80	90	Average
Methods	LR	0.01407 ±0.0007	0.01978 ±0.0099	0.02198 ±0.0086	0.01865 ±0.0068	0.02206 ±0.0075	0.02469 ±0.0059	0.03114 ±0.0032	0.02730 ±0.0036	0.02659 ±0.0053	—
	DTR	0.01253 ±0.0001	*0.01179 ±0.0020*	*0.01090 ±0.0027*	0.01257 ±0.0024	*0.01161 ±0.0035*	*0.01065 ±0.0042*	*0.00723 ±0.0001*	0.00978 ±0.0055	*0.00714 ±0.0002*	—
	CTR	0.01504 ±0.0001	0.01419 ±0.0013	0.01329 ±0.0021	0.01446 ±0.0020	0.01372 ±0.0030	0.01288 ±0.0036	0.00989 ±0.0002	0.01204 ±0.0051	0.00960 ±0.0004	—
	PMF	0.02091 ±0.0001	0.02083 ±0.0007	0.02035 ±0.0013	0.02129 ±0.0012	0.02136 ±0.0020	0.02067 ±0.0019	0.01848 ±0.0005	0.02018 ±0.0037	0.01918 ±0.0008	—
	NLMFGP	*0.01149 ±0.0001*	0.01527 ±0.0068	0.01359 ± 0.0037	*0.01221 ± 0.0013*	0.01210 ± 0.0022	0.01304 ± 0.0018	0.00774 ± 0.0008	*0.00926 ± 0.0057*	0.00738 ± 0.0016	—
	CPCTR	0.01507 ±0.0001	0.01404 ±0.0016	0.01316 ±0.0022	0.01426 ±0.0021	0.01355 ±0.0031	0.01260 ±0.0037	0.00929 ±0.0002	0.01176 ±0.0052	0.00908 ±0.0002	—
	GPCTR	**0.01144 ±0.0001**	**0.01103 ±0.0013**	**0.01007 ±0.0026**	**0.01158 ±0.0026**	**0.01058 ±0.0040**	**0.00963 ±0.0045**	**0.00572 ±0.0001**	**0.00856 ±0.0061**	**0.00549 ±0.0002**	—
Improvement of GPCTR vs.	LR(%)	18.69*	44.23	54.19*	37.91*	52.04*	61.00*	81.63*	68.64*	79.35*	55.30
	DTR(%)	8.70*	6.45	7.61	7.88	8.87*	9.58*	20.89*	12.47*	23.11*	11.73
	CTR(%)	23.94*	22.27*	24.23*	19.92*	22.89	25.23	42.16*	28.90*	42.81*	28.04
	PMF(%)	45.29*	47.05*	50.52*	45.61*	50.47*	53.41*	69.05*	57.58*	71.38*	54.48
	NLMFGP(%)	0.44	27.77*	25.90*	5.16	12.56	26.15*	26.10*	7.56	25.61*	17.47
	CPCTR(%)	24.09*	21.44*	23.48*	18.79	21.92	23.57	38.43*	27.21	39.54*	26.50

Note: The significant improvements of GPCTR over the comparative methods are marked by star superscript (Turkey's HSD, $p < .05$).



(a) Kanzhun. (b) Glassdoor.

Fig. 3. Salary prediction performance of CPCTR compared with its two variants. (a): Kanzhun; (b): Glassdoor.

average 4.24% relative RMSE improvement, while on Glassdoor GPCTR outperforms CPCTR by an average 26.50% relative RMSE improvement. This well demonstrates the great power of introducing GP in the model structure.

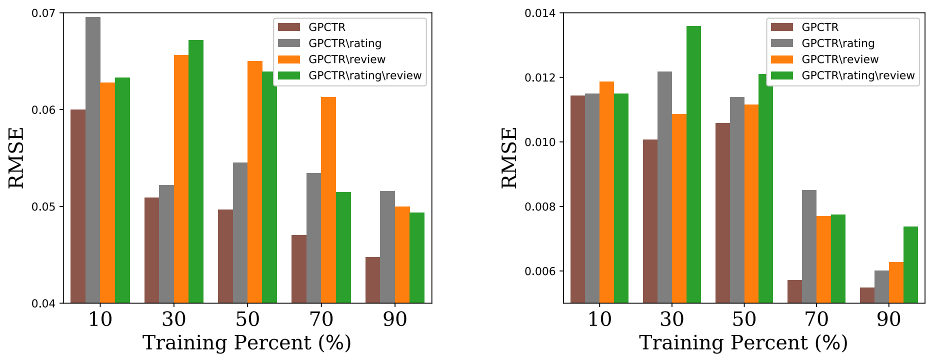
- Last, we can see that CTR shows unstable prediction performance compared to CPCTR or GPCTR, e.g., for all percentage of training set on Kanzhun, CTR shows poorer performance than PMF while for all percentage of training set on Glassdoor, CTR shows better performance than PMF. The reason might be that CTR only fuses textual opinion information for salary prediction and can neither utilize the rating information nor explicitly model the positive/negtive topic-word distributions, which makes it inaccurate to capture the intrinsic correlation of different data sources hidden inside latent spaces. Compared to the robust and stable performance of our proposed models, it further demonstrates the effectiveness of integrating three data sources in our framework.

5.4 Effects of Job-company Pair Modeling and Joint Optimization

To validate the effects of two key components in CPCTR, including job-company pair modeling and joint optimization, we design two variants of CPCTR as follows:

- **CPCTR-JobTheta** is a joint model of review ratings, texts, and salaries similar to CPCTR, but simply assumes that each job position rather than job-company pair has a topic proportion.
- **CPCTR-TwoStages** is also a joint model for incorporating review ratings, texts, and salaries, but follows a two-stage optimization procedure. That is to first employ topic models to learn topic proportion of each job-company pair, then conduct a joint factorization of ratings and salaries.

We empirically compare the two variants with CPCTR in the task of salary prediction. From the results in Figure 3, we can see that CPCTR-JobTheta cannot compete with our proposed CPCTR model on both Kanzhun and Glassdoor datasets. This well validates the effectiveness of job-company pair modeling in salary prediction for CPCTR. Moreover, it can be seen that CPCTR consistently outperforms CPCTR-TwoStages on Kanzhun and Glassdoor. This shows that the joint optimization of topic modeling and probabilistic matrix factorization indeed helps to predict salary more accurately.



(a) Kanzhun. (b) Glassdoor.

Fig. 4. Salary prediction performance of GPCTR compared with its three submodels. (a): Kanzhun; (b): Glassdoor.

5.5 Contribution Analysis of Data Sources from Reviews and Ratings

To study different data sources' contribution to our model, we remove the influence of review texts and ratings from GPCTR, respectively, and generate three submodels as follows:

- **GPCTR\rating**: Eliminating the contribution of ratings by removing the nonlinear matrix factorization over rating information in GPCTR.
- **GPCTR\review**: Eliminating the contribution of review texts by removing the probabilistic topic modeling of review texts in GPCTR.
- **GPCTR\rating\review**: Eliminating the contribution of both rating and review texts in GPCTR, which is equivalent to the baseline NLMFGP.

The performance of GPCTR and its three submodels for salary prediction are shown in Figure 4. It is obvious to see that the predictive performance becomes poorer when the impact of either review texts or ratings is removed from the model, which indicates that reviews and ratings are both essential for predicting salary information.

It is also interesting to see that the performance degradation of eliminating one single view of data source from GPCTR, e.g., only eliminating review texts or only eliminating ratings, is sometimes greater than that of eliminating both reviews and ratings. This suggests that incorporating only one view of side information into the framework might show unstable salary prediction performance, which greatly validates the necessity of integrating ratings and reviews simultaneously into the framework for accurate salary prediction.

5.6 Correlation between Ratings and Salaries

Revisit that both the proposed CPCTR model and the extension model GPCTR integrate reviews and ratings for salary prediction simultaneously. Here, we design experiments to evaluate the capabilities of the two models in capturing correlation patterns between numeric ratings and salaries. Specifically, we first train the models on the 80% training set of Kanzhun and Glassdoor, and then plot the correlation matrices between the predicted ratings and predicted salaries on the remaining 20% testing set in Figure 5. Left figure is the ground truth of the correlation between ratings and salaries on the test set, middle figure and right figure are the correlation between the predicted ratings and salaries for GPCTR and CPCTR, respectively. The correlation matrices are calculated by the frequency of occurrence of the two random variables, i.e., rating and salary range of each job-company.

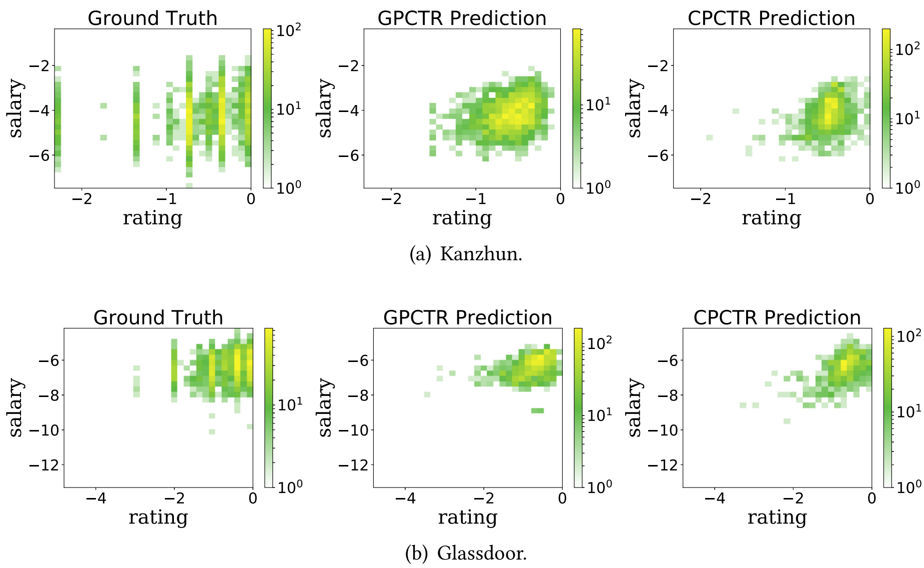


Fig. 5. Correlation matrices between ratings and salaries on testing set of (a) Kanzhun and (b) Glassdoor (percentage of training set = 80). Left: Ground truth; Middle: GPCTR; Right: CPCTR. Horizontal and vertical axes correspond to the $\log_2$ values of rating and salary, respectively.

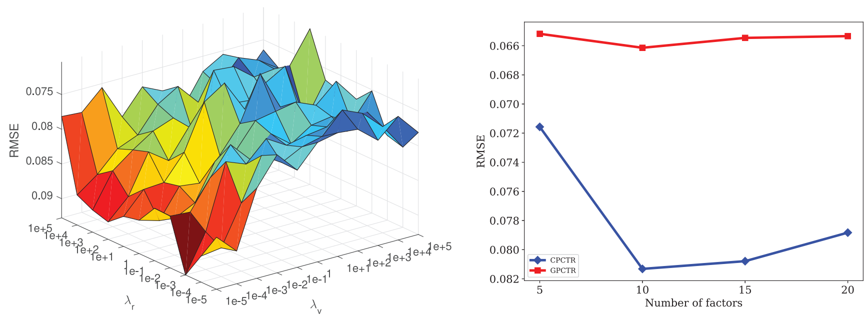
From the left subfigures of Figure 5, we can see that both the two datasets show complex nonlinear correlation between ratings and salaries. Comparing the predicted correlation plot of GPCTR and CPCTR, we can find that GPCTR captures more accurate correlation patterns than CPCTR, where GPCTR's correlation patterns are more alike to the ground truth distribution. Particularly on Glassdoor, it is obvious to see that the proposed CPCTR model tends to predict patterns, which shows positive correlation between ratings and salaries, whereas GPCTR approximates the shape of the true correlation distribution. These findings along with salary performance results induce the conclusion that GP can robustly capture the complex nonlinear correlation across different data views, which enhances our motivation of introducing GP in the model structure.

5.7 Parameter Sensitivity

In what follows, we study the effect of two parameters of CPCTR to salary prediction performance, i.e., λ_v, λ_r, where the content parameter λ_v controls the contribution of review content information to prediction performance and the rating parameter λ_r balances the contribution of rating information to model performance.

We fix the number of latent factors to $K = 5$, then vary λ_v and λ_r as from $1e-5$ to $1e+5$. The average performance within five times independent random run is shown in Figure 6(a). In general, the performance of CPCTR has been improved with increase of λ_v and λ_r, which validates the capabilities of review contents and ratings for enhancing salary prediction.

Next, we explore the impact of the number of latent factors K on salary prediction performance. As shown in Figure 6(b), we vary K in $\{5, 10, 15, 20\}$ and plot the corresponding performance of GPCTR and CPCTR, respectively. Generally, prediction performance of CPCTR decreases with increase of K, whereas GPCTR achieves relatively stable prediction performance with increase of K. This might be explained by CPCTR being prone to overfitting with large values of K. Also, it well validates the robustness of our extension GPCTR model with a more compact model structure.



(a) Impact of λ_v and λ_r. (b) Impact of K.

Fig. 6. Salary prediction performance with varying parameters. Note that all experiments are conducted on Kanzhun with percentage of training set = 80. (a): Impact of λ_v and λ_r for CPCTR. λ_v and λ_r range from $1e-5$ to $1e+5$, with number of latent factors fixed to $K = 5$. (b): Impact of K for GPCTR and CPCTR. Number of latent factors K ranges in $\{5, 10, 15, 20\}$, with other parameters fixed to the default.

Table 8. Statistics of Dataset for Product Brand Evaluation

Statistics	Amazon
# of Brands	5,977
# of Categories	10,474
# of Salaries/Ratings/Reviews	210,332
# of Pros Words	341,657,223
# of Cons Words	66,932,653
Matrix Sparsity	0.9966
Average Non-zero Entries per Brand	35.19
Average Pros Words per Category	32,619.56
Average Cons Words per Category	6,390.36

5.8 Evaluation on Model Generalizability

We have investigated the effectiveness and robustness of our proposed models in revealing useful and meaningful patterns of employer brands during evaluation through extensive experiments. Other than EBE, however, we further show in this section the generalizability of our models over areas like product brand evaluation.

In essence, our proposed framework can generally be applied for joint opinion mining and predictive modeling with a broad range of application scenarios. For example, in this popular Amazon review dataset[9] [He and McAuley 2016], users can give ratings and post review texts on the products, and products might have some metadata, including brands (e.g., Coxlures), categories (e.g., Sports & Outdoors) and prices. Similar to EBE, in this scenario, we can also exploit the advantages and disadvantages of product brands with reviews as well as quantitatively benchmark their competitiveness through product prices, which is termed as product brand evaluation.

We empirically evaluate the performance for price prediction on Amazon dataset, where the dataset statistics are shown in Table 8. Note that we take the same data preprocessing and evaluation protocols as those on online employment datasets such as Kanzhun and Glassdoor. It can be seen that the rating and price matrices are also sparse on Amazon; however, the average

[9] http://jmcauley.ucsd.edu/data/amazon/.

document length is much longer than that on Kanzhun/Glassdoor, e.g., the average number of pros words per category on Amazon is 141 times longer than the average number of pros words per job on Kanzhun. As such, Amazon has much richer textual information in terms of review texts than online employment sites like Kanzhun and Glassdoor. From the performance of price prediction in Table 9 along with the predictive performance on online employment datasets, we have the following observations:

- Our proposed models are generally applicable and competitive in different real-world application scenarios. For example, compared to PMF, GPCTR averagely gains 54.48% relative RMSE improvement on Glassdoor for salary prediction; while on Amazon price prediction task, the average value of relative RMSE improvement is 14.38%.
- It is also interesting to see that, compared to CPCTR, GPCTR gains an average 4.24%, 26.50% relative RMSE improvement on Kanzhun and Glassdoor, respectively; whereas on Amazon, the average relative improvement is only 1.66%. This might be owing to the richer textual information that Amazon dataset provides, and CPCTR exploits the semantics of textual review information more deeply, i.e., CPCTR models the topic-word distributions of each job while GPCTR models the topic-word distributions globally. This benefit in topic modeling for CPCTR might have an increasing influence to the predictive performance when the amount of review texts is large.

6 CONCLUSION

In this article, we addressed the Employer Brand Evaluation (EBE) problem where the heterogeneous textual information and numerical information of companies can be collaboratively modelled. We proposed CPCTR, a unified approach for learning the latent structural patterns of employer brands. A unique perspective of our approach is that it formulates a joint optimization framework for learning the latent patterns of employer brands, including the positive/negative opinions of companies and the latent topic variable that influences salary from an employee's perspective. With the identified patterns, both opinion analysis and salary benchmarking can be conducted effectively. Moreover, an extension model called GPCTR was also proposed based on GP to capture complex correlation across different data sources. Extensive experiments on real-world datasets demonstrated the effectiveness of CPCTR and GPCTR in leveraging heterogeneous company information for accurate salary benchmarking and comprehensive interpretation of employer competitiveness. The generalizability of our proposed models was also empirically evaluated over areas other than EBE.

APPENDIXES

A TECHNICAL DETAILS OF PARAMETER LEARNING

A.1 Detailed Derivation of Parameter Learning for CPCTR

Following generative process of CPCTR, we use capital letters (S, R, W^P, W^C) to represent the whole datasets of observed variables: $S = \{s_{j,e}\}_{j=1,e=1}^{J,E}$, $R = \{r_{j,e}\}_{j=1,e=1}^{J,E}$, $W^P = \{w_{n,j,e}^P\}_{n=1,j=1,e=1}^{N,J,E}$, $W^C = \{w_{m,j,e}^C\}_{m=1,j=1,e=1}^{M,J,E}$. Similarly, sets of parameters (latent variables) in the proposed model can be represented as: $\theta = \{\theta_{j,e}\}_{j=1,e=1}^{J,E}$, $\beta = \{\beta_{k,j}\}_{k=1,j=1}^{K,J}$, $\varphi = \{\varphi_{k,j}\}_{k=1,j=1}^{K,J}$, $U = \{u_j\}_{j=1}^{J}$, $B = \{b_j\}_{j=1}^{J}$, $V = \{v_{j,e}\}_{j=1,e=1}^{J,E}$.

Intuitively, parameter learning for the proposed model is quite straightforward: (1) $\lambda_\bullet = \{\lambda_u, \lambda_b, \lambda_v, \lambda_r, \lambda_s\}$ and α are treated as hyperparameters that are manually predefined and fixed throughout the entire learning process; (2) inference of (θ, U, V, B) and estimation of (β, φ) are conducted in an EM-like procedure, where (θ, U, V, B) are inferred with Maximum *a posteriori*

Table 9. Performance of Price Prediction on Amazon

Training (%)		10	20	30	40	50	60	70	80	90	Average
Methods	LR	0.22656 ±0.0041	0.24381 ±0.0485	0.11702 ±0.0007	0.12696 ±0.0284	0.08084 ±0.0001	0.07405 ±0.0008	0.07264 ±0.0052	0.06766 ±0.0005	0.06523 ±0.0007	—
	DTR	0.09075 ±0.0015	0.08996 ±0.0008	0.08779 ±0.0007	0.08799 ±0.0010	0.08703 ±0.0003	0.08603 ±0.0005	0.08641 ±0.0009	0.08639 ±0.0008	0.08682 ±0.0009	—
	CTR	0.08486 ±0.0001	0.08345 ±0.0002	0.08291 ±0.0004	0.08211 ±0.0002	0.08120 ±0.0006	0.07981 ±0.0005	0.07940 ±0.0005	0.07893 ±0.0008	0.07766 ±0.0007	—
	PMF	0.07963 ±0.0001	0.07935 ±0.0002	0.07978 ±0.0004	0.07983 ±0.0002	0.07983 ±0.0005	0.07920 ±0.0005	0.07961 ±0.0004	0.07979 ±0.0007	0.07924 ±0.0008	—
	NLMFGP	*0.07345 ± 0.0012*	0.07339 ± 0.0005	0.07388 ± 0.0004	0.07363 ± 0.0004	0.07384 ± 0.0007	0.07432 ± 0.0014	0.07312 ± 0.0013	0.07400 ± 0.0011	0.07170 ± 0.0012	—
	CPCTR	0.07626 ±0.0002	*0.07230 ±0.0003*	**0.07044 ±0.0002**	*0.06925 ±0.0003*	**0.06824 ±0.0002**	*0.06738 ±0.0006*	*0.06691 ±0.0005*	*0.06678 ±0.0008*	*0.06604 ±0.0010*	—
	GPCTR	**0.07332 ±0.0001**	**0.07179 ±0.0003**	*0.07052 ±0.0003*	**0.06911 ±0.0005**	*0.06835 ±0.0008*	**0.06668 ±0.0006**	**0.06526 ±0.0009**	**0.06485 ±0.0011**	**0.06337 ±0.0011**	—
Improvement of GPCTR vs.	LR(%)	67.64*	70.55*	39.74*	45.57*	15.45*	9.95*	10.16*	4.15*	2.85	29.56
	DTR(%)	19.21*	20.20	19.67*	21.46	21.46*	22.49*	24.48*	24.93*	27.01*	22.32
	CTR(%)	13.39*	13.97*	14.94*	15.83*	15.83*	16.45*	17.81*	17.84*	18.40*	16.05
	PMF(%)	7.92*	9.53*	11.61*	13.43*	14.38*	15.81*	18.03*	18.72*	20.03*	14.38
	NLMFGP(%)	0.18	2.18*	4.55*	6.14*	7.43*	10.28*	10.75*	12.36*	11.62*	7.28
	CPCTR(%)	3.86*	0.71*	-0.11	0.20	-0.16	1.04*	2.47*	2.89	4.04*	1.66

Note: The significant improvements of GPCTR over the comparative methods are marked by star superscript (Turkey's HSD, $p < .05$).

(MAP) estimates in E-step and (β, φ) are estimated by an empirical Bayes method in M-step [Blei et al. 2003; Wang and Blei 2011].

We first introduce the E-step. Maximization of the full posterior of (θ, U, V, B), given topic-word distribution parameters (β, φ) and other hyperparameters $(\lambda_\bullet, \alpha)$, is equivalent to maximizing the likelihood of complete data, including observed data (R, S, W^P, W^C) and latent variables (θ, U, V, B), given $(\lambda_\bullet, \beta, \varphi)$:

$$
\begin{aligned}
&p(\theta, U, V, B|S, R, W^P, W^C, \beta, \varphi, \lambda_\bullet, \alpha) \\
&\propto p(S, R, W^P, W^C, \theta, U, V, B|\beta, \varphi, \lambda_\bullet, \alpha) \\
&= p(U|\lambda_u)p(B|\lambda_b)p(V|\lambda_v, \theta)p(R, S|U, B, V, \lambda_r, \lambda_s)p(W^P, W^C|\beta, \varphi, \theta)p(\theta|\alpha) \\
&= \prod_j p(u_j|\lambda_u) \prod_j p(b_j|\lambda_b) \prod_e \prod_j p(v_{j,e}|\lambda_v, \theta_{j,e}) \prod_e \prod_j p(r_{j,e}|b_j, v_{j,e}, \lambda_r) \prod_e \prod_j p(s_{j,e}|u_j, v_{j,e}, \lambda_s) \\
&\quad \prod_e \prod_j \left(\prod_n p(w^P_{n,j,e}|\beta_j, \theta_{j,e}) \prod_m p(w^C_{m,j,e}|\varphi_j, \theta_{j,e}) \right) p(\theta|\alpha).
\end{aligned} \tag{18}
$$

We further calculate the logarithm of the above likelihood, forming the log likelihood of complete data as follows:

$$
\begin{aligned}
\mathcal{L} &= \log p(S, R, W^P, W^C, \theta, U, V, B|\beta, \varphi, \lambda_\bullet, \alpha) \\
&= -\frac{\lambda_b}{2} \sum_j b_j^\top b_j - \frac{\lambda_u}{2} \sum_j u_j^\top u_j - \frac{\lambda_r}{2} \sum_e \sum_j \left(r_{j,e} - b_j^\top v_{j,e} \right)^2 - \frac{\lambda_s}{2} \sum_e \sum_j \left(s_{j,e} - u_j^\top v_{j,e} \right)^2 \\
&\quad - \frac{\lambda_v}{2} \sum_e \sum_j (v_{j,e} - \theta_{j,e})^\top (v_{j,e} - \theta_{j,e}) + \sum_e \sum_j \left(\sum_n \log \left(\sum_k \theta_{k,j,e} \beta_{k, w^P_{n,j,e}} \right) \right. \\
&\quad \left. + \sum_m \log \left(\sum_k \theta_{k,j,e} \varphi_{k, w^C_{m,j,e}} \right) \right).
\end{aligned} \tag{19}
$$

We employ Coordinate Ascent (CA) approach to alternatively optimize the latent factors $\{u_j, b_j, v_{j,e}\}$ and the simplex variables $\theta_{j,e}$. For u_j, b_j, and $v_{j,e}$, we follow in a similar fashion as for basic matrix factorization [Hu et al. 2008]. Given the current estimation of $\theta_{j,e}$, taking the gradient of $\mathcal{L}$ with respect to u_j, b_j, $v_{j,e}$ and setting it to zero leads to

$$
u_j = \left(\lambda_u \mathbf{I}_K + \lambda_s \sum_e v_{j,e} v_{j,e}^\top \right)^{-1} \lambda_s \sum_e v_{j,e} s_{j,e}, \tag{20}
$$

$$
b_j = \left(\lambda_b \mathbf{I}_K + \lambda_r \sum_e v_{j,e} v_{j,e}^\top \right)^{-1} \lambda_r \sum_e v_{j,e} r_{j,e}, \tag{21}
$$

$$
v_{j,e} = (\lambda_v \mathbf{I}_K + \lambda_s u_j u_j^\top + \lambda_r b_j b_j^\top)^{-1} \left(\lambda_v \theta_{j,e} + \lambda_s u_j s_{j,e} + \lambda_r b_j r_{j,e} \right). \tag{22}
$$

However, for $\theta_{j,e}$, we cannot obtain its closed form by directly setting the gradient of $\mathcal{L}$ with respect to $\theta_{j,e}$ to zero due to the coupling between θ and β/φ. Thus, we resort to variational method for approximate inference. We first define $q(z^P_{n,j,e} = k) = \phi^P_{k,n,j,e}$ and $q(z^C_{m,j,e} = k) = \phi^C_{k,m,j,e}$. Then the items that contain $\theta_{j,e}$ are separated and Jensen's inequality is applied to obtain the likelihood bound:

$$
\mathcal{L}(\theta_{j,e}) \geq -\frac{\lambda_v}{2} (v_{j,e} - \theta_{j,e})^\top (v_{j,e} - \theta_{j,e}) + \sum_n \sum_k \phi^P_{k,n,j,e} (\log \theta_{k,j,e} \beta_{k, w^P_{n,j,e}} - \log \phi^P_{k,n,j,e})
$$

$$
\begin{aligned}
&+ \sum_m \sum_k \phi^C_{k,m,j,e} (\log \theta_{k,j,e} \varphi_{k,w^C_{m,j,e}} - \log \phi^C_{k,m,j,e}) \\
=& \mathcal{L}(\theta_{j,e}, \phi_{j,e}),
\end{aligned} \tag{23}
$$

where $\phi_{j,e} = \{\{\phi^P_{k,n,j,e}\}^{K,N}_{k=1,n=1}, \{\phi^C_{k,m,j,e}\}^{K,M}_{k=1,m=1}\}$. The optimal variational multinomial $\phi^P_{k,n,j,e}$ and $\phi^C_{k,m,j,e}$ satisfy

$$
\phi^P_{k,n,j,e} \propto \theta_{k,j,e} \beta_{k,w^P_{n,j,e}}, \tag{24}
$$

$$
\phi^C_{k,m,j,e} \propto \theta_{k,j,e} \varphi_{k,w^C_{m,j,e}}. \tag{25}
$$

The $\mathcal{L}(\theta_{j,e}, \phi_{j,e})$ gives a tight lower bound of $\mathcal{L}(\theta_{j,e})$. Similar to CTR [Wang and Blei 2011], we use projection gradient [Bertsekas 1999] to optimize $\theta_{j,e}$.

For M-step, following the same M-step of estimating topic-word distribution parameter in LDA [Blei et al. 2003], we optimize β and φ:

$$
\beta_{g,k,j} \propto \sum_e \sum_n \phi^P_{k,n,j,e} \mathbb{I}\left[w^P_{n,j,e} = g\right], \tag{26}
$$

$$
\varphi_{g,k,j} \propto \sum_e \sum_m \phi^C_{k,m,j,e} \mathbb{I}\left[w^C_{m,j,e} = g\right], \tag{27}
$$

where we denote g as an arbitrary term in the vocabulary set.

A.2 Detailed Derivation of Parameter Learning for GPCTR

For all E companies, the negative log likelihood on the observed set is as follows:

$$
\begin{aligned}
\mathcal{L} = \sum_{e=1}^{E} \Bigg[& \frac{N^r_e}{2} \log |C^r_e| + \frac{1}{2} r^\top_{o_e,e} C^{r-1}_e r_{o_e,e} + \frac{N^s_e}{2} \log |C^s_e| + \frac{1}{2} s^\top_{o_e,e} C^{s-1}_e s_{o_e,e} \\
& + \sum_{j \in o_e} \frac{\lambda_v}{2} (v_j - \theta_j)^\top (v_j - \theta_j) - \sum_{j \in o_e} \sum_n \log \left(\sum_{k=1}^{K} \theta_{j,k} \beta_{k,w^P_{j,n}} \right) - \sum_{j \in o_e} \sum_m \log \left(\sum_{k=1}^{K} \theta_{j,k} \varphi_{k,w^C_{j,m}} \right) \Bigg],
\end{aligned} \tag{28}
$$

where we defined the covariance function as

$$
\begin{aligned}
C^r_e &= \alpha^{-1}_r V_{o_e} V^\top_{o_e} + \lambda^{-1}_r \mathbf{I}, \\
C^s_e &= \alpha^{-1}_s V_{o_e} V^\top_{o_e} + \lambda^{-1}_s \mathbf{I},
\end{aligned}
$$

and N^r_e, N^s_e are the number of job positions rated by company e on the observed set of rating and salary matrix, respectively.

For inferring V, we aim at minimizing the objective function (namely, the negative log likelihood $\mathcal{L}$) with regard to V. Here, an SGD method is employed to optimize V. Specifically, at one epoch, for each company, we feed the observed ratings/salaries one at a time and compute the negative log likelihood's gradients for the company. Parameter V can then be updated by the gradients for that company. We compute the negative log likelihood for company e as follows:

$$
\begin{aligned}
\mathcal{L}_e(V) =& \frac{N^r_e}{2} \log |C^r_e| + \frac{1}{2} \left(r^\top_{o_e,e} C^{r-1}_e r_{o_e,e} \right) + \frac{N^s_e}{2} \log |C^s_e| \\
& + \frac{1}{2} (s^\top_{o_e,e} C^{s-1}_e s_{o_e,e}) + \sum_{j \in o_e} \frac{\lambda_v}{2} (v_j - \theta_j)^\top (v_j - \theta_j),
\end{aligned} \tag{29}
$$

and then compute the gradient with respect to V_{o_e} as

$$\frac{\partial \mathcal{L}_e(V)}{\partial V_{o_e}} = (-G_r - G_s)V_{o_e} + \lambda_v(V_{o_e} - \theta_{o_e}), \tag{30}$$

where we have defined

$$G_r = (C_e^{r-1} r_{o_e,e} r_{o_e,e}^\top C_e^{r-1} - C_e^{r-1}),$$
$$G_s = (C_e^{s-1} s_{o_e,e} s_{o_e,e}^\top C_e^{s-1} - C_e^{s-1}).$$

Given V, we now introduce the inference of the topic proportion vector θ_j. We introduce variational distributions $q(z_{n,j}^P = k) = \phi_{k,n,j}^P$, $q(z_{m,j}^C = k) = \phi_{k,m,j}^C$. The items that contain θ_j can then be separated and Jensen's inequality can be applied to obtain the likelihood bound as follows:

$$\begin{aligned}\mathcal{L}(\theta_j) \leq \frac{\lambda_v}{2}(v_j - \theta_j)^\top(v_j - \theta_j) - \sum_{n=1}^{N}\sum_{k=1}^{K} \phi_{k,n,j}^P(\log \theta_{k,j}\beta_{k,w_{n,j}^P} - \log \phi_{k,n,j}^P) \\ - \sum_{m=1}^{M}\sum_{k=1}^{K} \phi_{k,m,j}^C(\log \theta_{k,j}\varphi_{k,w_{m,j}^C} - \log \phi_{k,m,j}^C) \\ = \mathcal{L}(\theta_j, \phi_j),\end{aligned} \tag{31}$$

where $\phi_j = \{\{\phi_{k,n,j}^P\}_{k=1,n=1}^{K,N}, \{\phi_{k,m,j}^C\}_{k=1,m=1}^{K,M}\}$. In the E-step, we optimize the variational multinomial $\phi_{k,n,j}^P$ and $\phi_{k,m,j}^C$ as follows:

$$\phi_{k,n,j}^P \propto \theta_{k,j}\beta_{k,w_{n,j}^P}, \tag{32}$$

$$\phi_{k,m,j}^C \propto \theta_{k,j}\varphi_{k,w_{m,j}^C}. \tag{33}$$

Similar to CPCTR, projection gradient is utilized for optimizing θ_j. In the M-step, β and φ can be optimized as follows:

$$\beta_{g,k} \propto \sum_n \phi_{k,n,j}^P \mathbb{I}\left[w_{n,j}^P = g\right], \tag{34}$$

$$\varphi_{g,k} \propto \sum_m \phi_{k,m,j}^C \mathbb{I}\left[w_{m,j}^C = g\right], \tag{35}$$

where g is denoted as an arbitrary word in the vocabulary set.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their constructive comments on this article.

REFERENCES

Rakesh Kumar Agrawal and Pragati Swaroop. 2009. Effect of employer brand image on application intentions of B-school undergraduates. *Vision* 13, 3 (2009), 41–49.

Kristin Backhaus and Surinder Tikoo. 2004. Conceptualizing and researching employer branding. *Career Dev. Int.* 9, 5 (2004), 501–517. DOI: https://doi.org/10.1108/13620430410550754

Pierre Berthon, Michael Ewing, and Li Lian Hah. 2005. Captivating company: Dimensions of attractiveness in employer branding. *Int. J. Advert.* 24, 2 (2005), 151–172.

D. P. Bertsekas. 1999. *Nonlinear Programming.* Athena Scientific.

Shuqing Bian, Wayne Xin Zhao, Yang Song, Tao Zhang, and Ji-Rong Wen. 2019. Domain adaptation for person-job fit with transferable deep global match network. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* Association for Computational Linguistics, 4810–4820. DOI: https://doi.org/10.18653/v1/D19-1487

David M. Blei and John D. Lafferty. 2005. Correlated topic models. In *Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05).* The MIT Press, Cambridge, MA, 147–154. Retrieved from http://dl.acm.org/citation.cfm?id=2976248.2976267.

David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning (ICML'06).* ACM, New York, NY, 113–120. DOI: https://doi.org/10.1145/1143844.1143859

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3 (Mar. 2003), 993–1022. Retrieved from http://dl.acm.org/citation.cfm?id=944919.944937.

Edwin V. Bonilla, Shengbo Guo, and Scott Sanner. 2010. Gaussian process preference elicitation. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1 (NIPS'10).* Curran Associates Inc., 262–270. Retrieved from http://dl.acm.org/citation.cfm?id=2997189.2997219.

L. Breiman, J. H. Friedman, R. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees.* Wadsworth International Group.

Xi Chen, Yiqun Liu, Liang Zhang, and Krishnaram Kenthapadi. 2018. How LinkedIn economic graph bonds information and product: Applications in LinkedIn salary. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18).* ACM, New York, NY, 120–129. DOI: https://doi.org/10.1145/3219819.3219921

Neeti Leekha Chhabra and Sanjeev Sharma. 2014. Employer branding: Strategy for improving employer attractiveness. *Int. J. Organiz. Anal.* 22, 1 (2014), 48–60. DOI: https://doi.org/10.1108/IJOA-09-2011-0513

Martin R. Edwards. 2009. An integrative review of employer branding and OB theory. *Pers. Rev.* 39, 1 (2009), 5–23. DOI: https://doi.org/10.1108/00483481011012809

Carley Foster, Khanyapuss Punjaisri, and Ranis Cheng. 2010. Exploring the relationship between corporate, internal, and employer branding. *J. Prod. Brand Manag.* 19, 6 (2010), 401–409. DOI: https://doi.org/10.1108/10610421011085712

Francis Galton. 1886. Regression towards mediocrity in hereditary stature.*J. Anthropol. Inst. Great Brit. Irel.* 15 (1886), 246–263.

Gayatree Ganu, Noemie Elhadad, and Amélie Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In *Proceedings of the 12th International Workshop on the Web and Databases (WebDB'09)*, Vol. 9. 1–6. Retrieved from http://webdb09.cse.buffalo.edu/papers/Paper9/WebDB.pdf.

Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-aware ranking in search & recommendation systems with application to LinkedIn talent search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'19).* ACM, New York, NY, 2221–2231. DOI: https://doi.org/10.1145/3292500.3330691

Sahin Cem Geyik, Vijay Dialani, Meng Meng, and Ryan Smith. 2018a. In-session personalization for talent search. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM'18).* ACM, New York, NY, 2107–2115. DOI: https://doi.org/10.1145/3269206.3272012

Sahin Cem Geyik, Qi Guo, Bo Hu, Cagri Ozcaglar, Ketan Thakkar, Xianren Wu, and Krishnaram Kenthapadi. 2018b. Talent search and recommendation systems at LinkedIn: Practical challenges and lessons learned. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'18).* ACM, New York, NY, 1353–1354. DOI: https://doi.org/10.1145/3209978.3210205

G. H. Golub and C. Reinsch. 1970. Singular value decomposition and least squares solutions. *Numer. Math.* 14, 5 (Apr. 1970), 403–420. DOI: https://doi.org/10.1007/BF02163027

Viet Ha-Thuc, Yan Yan, Xianren Wu, Vijay Dialani, Abhishek Gupta, and Shakti Sinha. 2017. From query-by-keyword to query-by-example: LinkedIn talent search approach. In *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM'17).* ACM, New York, NY, 1737–1745. DOI: https://doi.org/10.1145/3132847.3132869

Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web (WWW'16).* International World Wide Web Conferences Steering Committee, 507–517. DOI: https://doi.org/10.1145/2872427.2883037

Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99).* ACM, New York, NY, 50–57. DOI: https://doi.org/10.1145/312624.312649

Neil Houlsby, Jose Miguel Hernández-Lobato, Ferenc Huszár, and Zoubin Ghahramani. 2012. Collaborative Gaussian processes for preference learning. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2.* Curran Associates Inc., 2096–2104. Retrieved from http://dl.acm.org/citation.cfm?id=2999325.2999369.

Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM'08).* IEEE Computer Society, Washington, DC, 263–272. DOI: https://doi.org/10.1109/ICDM.2008.22

Jack K. Ito, Céleste M. Brotheridge, and Kathie McFarland. 2013. Examining how preferences for employer branding attributes differ from entry to exit and how they relate to commitment, satisfaction, and retention. *Career Dev. Int.* 18, 7 (2013), 732–752. DOI: https://doi.org/10.1108/CDI-05-2013-0067

Krishnaram Kenthapadi, Stuart Ambler, Liang Zhang, and Deepak Agarwal. 2017. Bringing salary transparency to the world: Computing robust compensation insights via LinkedIn salary. In *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM'17).* ACM, New York, NY, 447–455. DOI: https://doi.org/10.1145/3132847.3132863

Pornthep Khongchai and Pokpong Songmuang. 2016. Implement of salary prediction system to improve student motivation using data mining technique. In *Proceedings of the 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS'16).* IEEE, 1–6.

Neil D. Lawrence and Raquel Urtasun. 2009. Non-linear matrix factorization with Gaussian processes. In *Proceedings of the 26th International Conference on Machine Learning (ICML'09).* ACM, New York, NY, 601–608. DOI: https://doi.org/10.1145/1553374.1553452

A Lazar. 2004. Income prediction via support vector machine. In *Proceedings of the International Conference on Machine Learning and Applications.* IEEE, 143–149.

Xiaopeng Li and James She. 2017. Collaborative variational autoencoder for recommender systems. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17).* ACM, New York, NY, 305–314. DOI: https://doi.org/10.1145/3097983.3098077

Filip Lievens and Scott Highhouse. 2003. The relation of instrumental and symbolic attributes to a company's attractiveness as an employer. *Personn. Psychol.* 56, 1 (2003), 75–102.

Filip Lievens, Greet Van Hoye, and Frederik Anseel. 2007. Organizational identity and employer image: Towards a unifying framework. *Brit. J. Manag.* 18 (2007), S45–S59.

Hao Lin, Hengshu Zhu, Yuan Zuo, Chen Zhu, Junjie Wu, and Hui Xiong. 2017. Collaborative company profiling: Insights from an employee's perspective. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI'17).* AAAI Press, 1417–1423. Retrieved from http://dl.acm.org/citation.cfm?id=3298239.3298446.

Chenghao Liu, Tao Jin, Steven C. Hoi, Peilin Zhao, and Jianling Sun. 2017. Collaborative topic regression for online recommender systems: An online and Bayesian approach. *Mach. Learn.* 106, 5 (May 2017), 651–670. DOI: https://doi.org/10.1007/s10994-016-5599-z

Benjamin Marlin. 2003. Modeling user rating profiles for collaborative filtering. In *Proceedings of the 16th International Conference on Neural Information Processing Systems (NIPS'03).* The MIT Press, Cambridge, MA, 627–634. Retrieved from http://dl.acm.org/citation.cfm?id=2981345.2981424.

Benjamin Marlin. 2004. *Collaborative Filtering: A Machine Learning Perspective.* University of Toronto.

Benjamin Marlin and Richard S. Zemel. 2004. The multiple multiplicative factor model for collaborative filtering. In *Proceedings of the 21st International Conference on Machine Learning (ICML'04).* ACM, New York, NY, 73–. DOI: https://doi.org/10.1145/1015330.1015437

Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys'13).* ACM, New York, NY, 165–172. DOI: https://doi.org/10.1145/2507157.2507163

Q. Meng, H. Zhu, K. Xiao, and H. Xiong. 2018. Intelligent salary benchmarking for talent recruitment: A holistic matrix factorization approach. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'18).* 337–346.

Mamta Mohapatra and Priyanka Sahu. 2018. Building a sustainable talent acquisition model in a dynamic business environment. *Int. J. Hum. Cap. Inf. Technol. Prof.* 9, 3 (July 2018), 42–52. DOI: https://doi.org/10.4018/IJHCITP.2018070103

Julie N. Cho. 2017. Contribution of talent analytics in change management within project management organizations the case of the French aerospace sector. *Proc. Comput. Sci.* 121, C (Jan. 2017), 625–629. DOI: https://doi.org/10.1016/j.procs.2017.11.082

Cagri Ozcaglar, Sahin Geyik, Brian Schmitz, Prakhar Sharma, Alex Shelkovnykov, Yiming Ma, and Erik Buchanan. 2019. Entity personalized talent search models with tree interaction features. In *Proceedings of the the World Wide Web Conference (WWW'19).* ACM, New York, NY, 3116–3122. DOI: https://doi.org/10.1145/3308558.3313672

Sanjay Purushotham, Yan Liu, and C.-C. Jay Kuo. 2012. Collaborative topic regression with social matrix factorization for recommendation systems. In *Proceedings of the 29th International Conference on International Conference on Machine Learning (ICML'12).* Omnipress, Madison, WI, 691–698.

Chuan Qin, Hengshu Zhu, Tong Xu, Chen Zhu, Liang Jiang, Enhong Chen, and Hui Xiong. 2018. Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'18)*. ACM, New York, NY, 25–34. DOI: https://doi.org/10.1145/3209978.3210025

Chuan Qin, Hengshu Zhu, Tong Xu, Chen Zhu, Chao Ma, Enhong Chen, and Hui Xiong. 2020. An enhanced neural network approach to person-job fit in talent recruitment. *ACM Trans. Inf. Syst.* 38, 2 (Feb. 2020). DOI: https://doi.org/10.1145/3376927

Rohan Ramanath, Hakan Inan, Gungor Polatkan, Bo Hu, Qi Guo, Cagri Ozcaglar, Xianren Wu, Krishnaram Kenthapadi, and Sahin Cem Geyik. 2018. Towards deep and representation learning for talent search at LinkedIn. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM'18)*. ACM, New York, NY, 2253–2261. DOI: https://doi.org/10.1145/3269206.3272030

Carl Edward Rasmussen and Christopher K. I. Williams. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI'04)*. AUAI Press, Arlington, Virginia, 487–494. Retrieved from http://dl.acm.org/citation.cfm?id=1036843.1036902.

Tetsuya Sakai. 2018. *Laboratory Experiments in Information Retrieval*. Vol. 40. Springer.

Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic matrix factorization. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07)*. Curran Associates Inc., 1257–1264. Retrieved from http://dl.acm.org/citation.cfm?id=2981562.2981720.

Ying Sun, Fuzhen Zhuang, Hengshu Zhu, Xin Song, Qing He, and Hui Xiong. 2019. The impact of person-organization fit on talent management: A structure-aware convolutional neural network approach. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'19)*. ACM, New York, NY, 1625–1633. DOI: https://doi.org/10.1145/3292500.3330849

Ivan Titov and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the ACL-08: HLT*. Association for Computational Linguistics, 308–316. Retrieved from http://www.aclweb.org/anthology/P/P08/P08-1036.

K. Vivekanandan and J. Soonu Aravindan. 2014. Aspect-based opinion mining: A survey. *Int. J. Comput. Appl.* 106, 3 (2014), 21–26.

Chong Wang and David M. Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*. ACM, New York, NY, 448–456. DOI: https://doi.org/10.1145/2020408.2020480

Hao Wang, Binyi Chen, and Wu-Jun Li. 2013. Collaborative topic regression with social regularization for tag recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*. AAAI Press, 2719–2725.

Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*. ACM, New York, NY, 1235–1244. DOI: https://doi.org/10.1145/2783258.2783273

Chunyan Xie, Richard P. Bagozzi, and Kjersti V. Meland. 2015. The impact of reputation and identity congruence on employer brand attractiveness. *Mark. Intell. Plan.* 33, 2 (2015), 124–146. DOI: https://doi.org/10.1108/MIP-03-2014-0051

Huang Xu, Zhiwen Yu, Jingyuan Yang, Hui Xiong, and Hengshu Zhu. 2018. Dynamic talent flow analysis with deep sequence prediction modeling. *IEEE Trans. Knowle. Data Eng.* 31, 10 (2018), 1926–1939.

Rui Yan, Ran Le, Yang Song, Tao Zhang, Xiangliang Zhang, and Dongyan Zhao. 2019. Interview choice reveals your preference on the market: To improve job-resume matching through profiling memories. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'19)*. ACM, New York, NY, 914–922. DOI: https://doi.org/10.1145/3292500.3330963

Guangxiang Zeng, Hengshu Zhu, Qi Liu, Ping Luo, Enhong Chen, and Tong Zhang. 2015. Matrix factorization with scale-invariant parameters. In *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*. AAAI Press, 4017–4024. Retrieved from http://dl.acm.org/citation.cfm?id=2832747.2832809.

Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. Association for Computing Machinery, New York, NY, 353–362. DOI: https://doi.org/10.1145/2939672.2939673

Le Zhang, Hengshu Zhu, Tong Xu, Chen Zhu, Chuan Qin, Hui Xiong, and Enhong Chen. 2019. Large-scale talent flow forecast with dynamic latent factor model? In *Proceedings of the World Wide Web Conference (WWW'19)*. Association for Computing Machinery, New York, NY, 2312–2322. DOI: https://doi.org/10.1145/3308558.3313525

Chen Zhu, Hengshu Zhu, Yong Ge, Enhong Chen, and Qi Liu. 2014. Tracking the evolution of social emotions: A time-aware topic modeling perspective. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'14)*. IEEE Computer Society, Washington, DC, 697–706. DOI: https://doi.org/10.1109/ICDM.2014.121

Chen Zhu, Hengshu Zhu, Hui Xiong, Chao Ma, Fang Xie, Pengliang Ding, and Pan Li. 2018. Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Trans. Manag. Inf. Syst.* 9, 3 (Sept. 2018). DOI: https://doi.org/10.1145/3234465

Received August 2019; revised April 2020; accepted April 2020