International Journal of Management, IT & Engineering
Vol. 14 Issue 04, April 2024
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

A framework for Labor Market Analysis using Machine Learning

Kumar Shanu
Student, MCA
School of Computer Application, Lovely Professional University, Punjab

Dr. Raj Sinha
Assistant Professor School of Computer Applications
Lovely Professional University, Punjab

Abstract: As professional workers are important in state-of-the-art information-based totally financial system, opposition for expertise among corporations is fierce. This paper proposes the usage of full-size quantities of textual content statistics to expect future competition inside the chinese language labor marketplace. present strategies rely on studying product and talent overlap among corporations. right here, we introduce two new metrics: sentiment evaluation of online reviews to gauge worker satisfaction, and a combined analysis thinking about both talent and sentiment. Our results show those new metrics outperform traditional methods, offering valuable insights for groups navigating an aggressive process market.

Keywords: Workers, Sentiment Analysis, Employee Satisfaction

Introduction

The demand for professional labor has accelerated dramatically in the cutting-edge knowl- side-based economic system, that's characterised by the developing depth in exertions marketplace competition among firms. however, the fast adjustments in science and generation make it extra difficult for firms to properly teach and expand employees to fulfill companies' needs for talent by means of themselves, and they increasingly rely upon the acquisition of human belongings from different companies to meet their human capital needs (Cappelli 2008; Lee, Mauer, and Xu 2018). consequently, it would be of special interest to identify future exertions marketplace competitors, that's essential for all firms' future strategic development (Friesenbichler and Reinstaller 2021).

At present, the good-sized quantity modern textual records are unexpectedly increasing. Liu, Pant, and Sheng (2020) noticed the opportunity for identifying labor market competitors in mild modern the increasing availability modern day textual statistics and latest advances in big records analytics, and used the textual facts to con- struct the human capital overlap and product overlap metrics as predictors to predict the labor market opposition within the america. primarily based on these metrics, this paper tries to experiment with device gaining knowledge state modern techniques to are expecting chinese hard work market opposition with chinese textual content statistics.

Moreover, sentiment evaluation has come to be famous recently (Bandhakavi et al. 2021; Osmani, Mohasefi, and Gharehchopogh 2020; Rehman et al. 2021; Yan et al. 2021), which has been broadly used in online tourism evaluations (Luo et al. 2021), product opinions (Al- Sharuee, Liu, and Pratama 2021; Vo, Nguyen, and very well2020), stock marketplace facts (Yin, Wu, and Kong 2020) and so on. but, trendy loss of information, there are few existing studies the usage of sentiment evaluation method cutting-edge firms' online critiques. In fact, Maimai's platform can offer a rich facts basis for these studies, in which personnel can submit their critiques approximately companies. those evaluations might also have an effect on task pleasure state-of-the-art employees and whether or not the indivi- twin works there. In reaction to this studies hole, this take a look at takes the human capital overlap and product overlap metrics as the fundamental metrics, and is a first try to construct the sentiment analysis metric via mining the emotional content expressed inside the companies' on-line opinions on Maimai's platform, that could provide the splendid predictive strength for identifying exertions market competition and deliver deep insights into labor market opposition.

The contributions cutting-edge this paper is supplied as follows:

- based totally on the human capital overlap and product overlap metrics, we endorse the sentiment analysis metric as a predictor to predict chinese language exertions market opposition by way of making use of guide Vector Machines (SVM) for binary sentiment class present day corporations' on-line reviews on Maimai's platform.
- We are expecting Chinese language labor market place opposition experimented with contemporary 9a2d564f1275e41c4e633abc33154 7db gadget contemporary strategies by way of the use of those metrics.
- We carry out the two- dimensional (second) competition analysis modern-day common industries to provide a greater nuanced photo ultra-modern the chinese language enterprise-unique competitor panorama.

The remainder latest this paper is organized as follows. phase 2 offers a brief overview contemporary the related work. section three describes the datasets and predictors. segment four affords empirical outcomes today's the prediction models and plays the nuanced 2d competition analysis. section 5 gives the conclusions.

Related Work

extensive literature investigated hard work market opposition primarily based at the seasoned- duct overlap and human capital overlap from special perspectives. in this component, the literature on conventional methods for measuring product overlap and human capital overlap is reviewed, respectively. subsequent, the paper that applied sentiment evaluation procedures is likewise reviewed.

The Measures of Product Overlap

The product overlap of companies is a basic predictor of their hard work market competition, and the measure of product overlap has attracted more interest in the subject of competitor evaluation. Hou and Robinson (2006) and Giroud and Mueller (2011) used the Herfindahl Index to measure product market competition. despite the fact that, this degree only represents the overlap within a product marketplace or maybe within an enterprise, it has less potential to represent both inside and across industry similarity in product marketplace and typically has much stronger records requirements.

Then, a few research see the opportunity for a brand new and extra trendy methodology in mild of the growing availability of public, unstructured data and latest advances in large information analytics (Java et al. 2007). To capture the associated- ness of firms within the product market space, Hoberg and Phillips (2010) evaluated the similarity of the textual content in product descriptions. Li, Lundholm, and Minnis (2013) implemented a competition measure inside the product marketplace derived from the textual analysis of corporations' 10-ok filings by rating the frequency of competition- associated phrases every yr and putting it into deciles. yet, they only can degree the average opposition pressure confronted by way of the available companies.

Pant and Sheng (2015) used firms' web sites and links to measure the product similarity and predict product marketplace competitors. a few industrial company profiling groups along with Hoover's and Mergent manually diagnosed product market competition. however, because of the size and dynamic nature of the problem, the strategies require extra manual preprocessing and may increase the complexity of representing the enterprise surroundings. normal, the previous measures are quite incomplete and do no longer explicitly cowl the product and marketplace areas when figuring out competitors. Shi, Lee, and Whinston (2016) proposed a brand-new statistics-analytic method to degree corporations' product overlap primarily based on industry classifications and commercial enterprise descriptions. specifically, they analyzed the unstructured texts that describe corporations' organizations using the statistical mastering approach of subject matter modeling, and built a novel business proximity degree based at the output. while compared with other strategies, this approach is scalable for big datasets and presents finer granularity on quantifying firms' positions inside the areas of product, market, and era. consequently, following Shi, Lee, and Whinston (2016), this paper constructs the product overlap metrics because the predictors in phase three.2.1 to discover chinese hard work marketplace competitors.

The Measures of Human Capital Overlap

The human capital overlap of companies is an important predictor of their hard work marketplace competition, and a challenging hassle that arises in identifying labor marketplace competitors is the measure of human capital overlap at the extent of a firm. Leping (2009) built a talent-based totally measure for human capital specificity and the degree is based totally on the opportunities of

utilizing competencies at the exertions marketplace, which depends on the quantity of jobs in which any precise talent is needed in labor marketplace. Dallimore (2010) proposed an inclusive degree of human capital by integrating conventional measures of human capital and non-accounting measures (e.g., educational degree, enjoy, and motivation). Santarelli and Thu (2013) selected individual-stage professional training, begin-up revel in, and mastering to measure human capital, the first dimensions of human capital are measured with traditional indicators and that they defined mastering because the capacity to accumulate expertise to conduct innovation sports (new product introduction, product innovation, and procedure innovation). Hun, Mauer, and Qianying (2018) received an industry occupation profile vector with elements same to the proportion of overall employment within the enterprise's occupations to degree human capital similarities amongst different firms. Sasso and Ritzen (2019) investigated human capital measures by using sectoral capabilities that described because the average cognitive abilities of the work- pressure in every usa- area mixture. Yang (2020) used a greater state-of-the-art measure of the human capital, in addition to education, which specializes in the competencies derived from self- assessed investments in task- specific human capital.

but, maximum of the studies on this discipline is aimed toward measuring the explicit understanding and talents, the implicit knowledge (furnish 1996) held through employees stays an open hassle within the human capital degree, which is likewise a critical component of human capital. Liu, Pant, and Sheng (2020) addressed those problems of human capital degree with a unique longitudinal business enterprise- worker matched statistics set from the information of online customers' public profile pages. They proposed novel human capital overlap metrics based totally on companies' ability endowment and their embedded human capital flow (HCF) network structure, which could capture the interfirm similarities inside the express understanding and tacit understanding base, respectively. consequently, the metrics could be applied more commonly to empirically test other human capital and company method theories and derive commercial enterprise intelligence (Newbert 2007). As a result, following Liu, Pant, and Sheng (2020), this paper constructs the human capital overlap metrics because the predictors in section.

Sentiment Analysis

Sentiment analysis can be used to help us attain emotional facts by means of mining and analyzing the emotional content expressed and aim to are expecting the orientation of sentiment gift at the massive textual data (Abbasi, Chen, and Salem 2008). This form of analysis is especially divided into dictionary-primarily based sentiment analysis and device getting to know-primarily based sentiment analysis.

Dictionary-Based Sentiment Analysis

Taboada et al. (2011) presented a lexicon-based totally technique to extracting sentiment from text. Hogenboom et al. (2014) explored the enlargement of lexicon-based totally sentiment analysis from English to Dutch, and they created the language-precise lexicon thru semantics. Sharma and Dutta (2021) applied Lexicon-based totally techniques, which used sentiment orientation rankings of phrases contained in the text for polarity willpower of documents. But, dictionary-based totally sentiment evaluation is limited through the richness of context and semantic expression, and those constraints typically make the accuracy price low.

Machine Learning-Based Sentiment Analysis

Sentiment analysis strategies based on extraordinary gadget getting to know algorithms are used for sentiment category in the current studies (Mohammadi and Shaverizade 2021; Indrawan et al. 2020). Shyamasundar and Jhansi (2020) proposed a multi-tier architecture for sentiment category, the feelings of large number of tweets generated from Twitter had been analyzed the use of device learning algorithms. in addition, Saura, Palacios-Marqués, and Ribeiro- Soriano (2022a) carried out machine mastering algorithms for sentiment evaluation of tweets. In some other observe, Textblob worked with system gaining knowledge of to perform sentiment analysis (Saura, Ribeiro-Soriano, and Iturricha-Fernández 2022b).

The SVM is appeared as one of the most effective gadgets mastering algorithms for sentiment classification (Bogawar and Bhoyar 2018). furthermore, a number of works have proven that with the aid of using SVM, no longer only the class performance can be stepped forward, but additionally the Vapnik–Chervonenkis (VC) dimension can be reduced (Dangi, Bhagat, and Dixit 2021; Liu, Bi, and Fan 2017; Na, Khoo, and Wu 2005). In precis, the SVM can offer accurate technical guide for research in sentiment type. consequently, this paper proposes our sentiment analysis metric as a predictor to become aware of chinese language labor market competition by way of making use of SVM in phase 3.2 three.

Datasets and Predictors

Datasets

We seeded the information through specializing in all China’s A- proportion listed firms (four,616 firms in general, along with all A- share indexed firms in Shanghai and Shenzhen inventory Markets in China) in the CSMAR Database (China stock market & Accounting research Database) as of December 2020. One purpose for choosing these indexed corporations as our seeds is that they may be worried with diverse industries in China Mainland, therefore leading to a variety of firms in our statistics set. in addition, due to the fact these listed corporations constitute the most valued firms via the marketplace, their employees on common might tend to symbolize the extra extraordinarily valued human capital. Maimai is a China-based totally profession and social-networking platform, it now has tens of millions of customers and is the most used professional social networking website in China. This offers us the threat to pick 3,0.5 China’s A- proportion listed firms that have extra than a hundred employees with actual-name registrations and entire public profiles on Maimai’s platform. Then, we crawled a complete of 50,508 public profiles of those employees, who're monthly active customers and more likely to be the goal of labor market competition between firms. The worker’s profile incorporates job stories, profession tags, and schooling statistics. in particular, the process experience of a character includes the firm name and the start and stop dates of this task experience. for this reason, from a character’s process studies, we will take a look at where the individual turned into operating (a number of the three,1/2 corporations protected in us examine) in a selected year. As a end result, a total of seventy eight,027 supply-target- year firm pairs from 2006 to 2020 are protected in our evaluation. The profession tags of an individual report a set of talent terms to signify the human capital the man or woman possess. we will mixture these person’s ability terms on the company stage. furthermore, we are able to construct a skill precis for each company. desk 1 suggests the top-10 abilities for some of the companies in specific industries in 2019, in which every ability term is weighted through the number of personnel on the given firm that mentioned it. The training data reports the man or woman’s university or better stage of training facts.

in addition to seeding our records from a diverse set of corporations, we carry out robustness assessments to affirm the representativeness of our records, as defined in Appendix A. firstly, we affirm that our data includes companies throughout all major enterprise organizations, and their distribution across industries is much like all China’s A-share listed companies covered in CSMAR Database. Secondly, we test.

Table 1. Top-10 skills of example firms in 2019.

ZTE Corporation	Vanke	SUNING	Ping An Bank
Java (85)	Project management (103) Retail business (91)		Big data (95)
Python (75)	Communication skills (98) E-commerce (89)		Finance (92)
IT development (69)	Negotiation ability (97) B usiness management		Performance management
	(78)		(89)
Excel (64)	Real estate development Br and strategy (72) (95)		Third-Party payment (85)
Cloud computing (52)	Architecture design (83)	Marketing (70)	Execution (82)
Communication R&D (51)	Teamwork (80)	Self-driven (65)	Credit business (78)
Internet testing (47)	Marketing planning (75)	Execution (65)	Risk management (74)
Equipment manuf	Project planning (69)	Internet op	Stress resistance (67)

how our sample matches with all employees at exceptional companies over time to confirm that our records are an inexpensive illustration of corporations in terms in their relative sizes. in the end, based totally on worker skills and business summaries of companies in our records, we display that the abilities of the personnel in our sample information are true representatives of the enterprise activities in their respective corporations.

Predictors

The product overlap metrics, human capital overlap metrics, sentiment ana- lysis metric and primary financial metrics of a couple of companies can be anticipated to be the predictors in their exertions market competition.

Product Overlap Metrics

The product overlap among companies consists of enterprise Code similarity, the similarity among companies' enterprise scopes, commercial enterprise scope topics, predominant organizations, and predominant business subjects. according to industry Code type (The steering for enterprise category of indexed organizations released through China Securities Regulatory fee in 2012) and following Shi, Lee, and Whinston (2016), we define industry Code similarity (Industry Code Sim) among a pair of corporations as proven in table 2.

Table 2. Metric for measuring Industry Code similarity (*IndustryCodeSim*) between firms.

Type	Example	Similarity Score
Digit 1 different	C14 (Food manufacturing) and A03 (Animal husbandry)	0
Digit 1 and 2 same, digit 3 Different	C14 (Food manufacturing) and C27 (Pharmaceutical manufacturing)	1
Same three digits	C14 (Food manufacturing) and C15(Beverage and refined tea manufacturing)	2
different	C14 (Food manufacturing) and C14 (Food manufacturing)	3

Note, C: Manufacturing; A: Agriculture, forestry, animal husbandry, and fishing.

Manifestly, the industry Code of a firm might not mirror all of the specific and granular product areas in which the firm operates. Then, we also compute company pair not handiest cosine similarity in the textual content terms of their enterprise scopes (BusscoTermSim) and business scope topics1 (BusscoTopicSim), however additionally cosine similarity inside the text terms of their principal groups (MainbusTermSim) and most important business topics2 (MainbusTopicSim) as the complementary metrics. The results are proven in figure B1, discern B2, discern B3 and determine B4 in Appendix B, respectively.

Human Capital Overlap Metrics

And there are commonly two factors of human capital overlap– exertions overlap and HCF community overlap (Liu, Pant, and Sheng 2020). hard work overlap consists of ability-time period similarity and talent topic similarity, that can seize the interfirm similarities within the express know-how base.

firstly, we define the talent-term similarity. In our facts, 50,508 employees suggested their 5,022 awesome skill phrases of their profiles' career tags, that is the idea for the construction of interfirm ability-primarily based similarity metrics (the talent phrases which are reported with the aid of a couple of workers are protected). As defined in section three.1, we constructed a talent precis for every company via aggregating the skill terms of its personnel in a particular 12 months (see table 1). For company ok, we denote sk in RN space as its ability vectorin which N five,022 is the set of all skill phrases across personnel in our facts. moreover, SF IFFs; ok stands for each element of sk, which may be calculated as follows:

SF IFFs;*k* = *SFs*;*k* × *IFFs*

In which the talent frequency SFs;k is defined because the range of personnel at firm k that reported the talent term s of their profiles. The inverse company frequency IFFs is defined as log F, where F is the total wide variety of corporations, and FFs is the quantity of firms whose talent precis incorporates the talent time periods.

primarily based on this formular, we measure skill-time period similarity (SkillTermSim) in the human capital at firms by means of the cosine similarity between the talent vectors similar to companies x and y as follows:

sim x y

(;) =

$\| sx \| \cdot \| sy \|$

Secondly, we outline the ability subject matter similarity. We follow Latent Dirichlet Allocation (LDA) (Blei 2012; Blei, Ng, and Jordan 2003) to discover the ability subjects inside the personnel' ability phrases. to use the LDA algorithm, we need firstly decide the wide variety of latent topics. primarily based on discern C1 in Appendix C, we set the wide variety of talent topics as 9 as it affords semantically significant subjects. Then, the LDA set of rules represents each of the nine talent subjects with a distribution over the five,022 talent terms (The pinnacle- 10 talent terms

$\underline{\quad s_x \cdot s_y \quad}$

with the best chances for every of the 9 talent topics are proven in figure C2 in Appendix C). therefore, each employee in our statistics may be represented by way of 9 chance values similar to the 9 ability subjects and each company okay in a year can be represented by way of a vector θokay of size 9, in which each element of the vector is the sum of its employees' probabilities for that skill topic. eventually, we calculate skill subject matter similarity (SkillTopicSim) within the human capital at corporations by the cosine similarity among the skill topic vectors similar to firms x and y as follows

sim x y θx · θy

‖ θx ‖ · ‖ θy ‖

The HCF network3 overlap includes the upstream similarity and downstream similarity, that can provide cues to degree the human capital overlap in terms of tacit expertise (Liu, Pant, and Sheng 2020). first of all, we define the upstream similarity. For firm k, we denote uk in RM space as its upstream vector, where M is the set of all firms' nodes in the HCF community. The detail uik of uk is the variety of personnel who've moved inside the past from company i to company ok. consequently, United Kingdom represents the distribution of employees who've migrated to firm okay over all firms. Then, we outline the upstream similarity (UpstreamSim) between companies x and y by means of the cosine similarity between the upstream vector ux and uy similar to the corporations as follows:

sim x y ux · uy ‖ ux ‖ · ‖ uy

further, the downstream vector dk stands for the distribution of personnel who have moved from company ok to other firms, and we can also denote the downstream similarity (DownstreamSim) among a couple of corporations as:

sim x y dx · dy ‖ dx ‖ · ‖ dy

Sentiment Analysis Metric

Sentiment evaluation has emerged as popular these days, which has been extensively utilized in online tourism reviews (Luo et al. 2021), product critiques (Al-Sharuee, Liu, and Pratama 2021; Vo, Nguyen, and very well2020), stock market information (Yin, Wu, and Kong 2020) and so on. however, due to loss of information, studies on sentiment evaluation of the corporations' on-line critiques are few to become aware of future hard work marketplace competition. In truth, Maimai's platform can provide a rich statistics basis for this research, where employees can put up their critiques approximately firms. those reviews of corporations are mainly about the wages, working hours, autonomy given to personnel, organizational structure and verbal exchange among employees and management, which might also affect process delight of employees and whether or not the character works there. For predicting exertions market competition, we crawled a complete of 5,125 on-line critiques of indexed firms on Maimai's platform from 2006 to 2020 and construct our sentiment analysis metric, which may provide the outstanding predictive software for identifying hard work marketplace competition and supply deep insights into exertions marketplace opposition. (several critiques of listed companies are supplied as examples in Appendix D).

on this paper, we use Python program to assemble the sentiment classification model. The specific technique is as follows:

This takes a look at uses Python and Jieba chinese phrase segmentation library to finish the processing of the pattern textual content-based records. all through word segmentation, we to begin with dispose of all punctuations and various symbols via regularity, gain pure text, and then load the forestall phrase library and self- built vocabulary for word segmentation. We pick the stop phrases used by the contemporary version of the NLPIR participle of the chinese language Academy of Sciences. simultaneously, we construct a specific lexicon on the idea of the prevailing glossary to prevent the department of names of companies and different phrases with special characteristics in this field.

The SVM model schooling only supports numerical samples; therefore, the sample textual content statistics have to be quantized. Word2Vec is extensively utilized in natural language processing obligations (e.g., textual content sentiment analysis) as the basic era in the subject of herbal language processing (Zhang et al. 2015). We use the Wiki chinese language corpus because the unique sample and choose the Word2Vec version of the Gensim library for training and CHI rectangular approach as a feature extraction method.

(three) This takes a look at manually marks 1,000 positive on-line opinions and 500 poor on line critiques because the training set. To achieve the proper model, we use Python's scikit-examine library, constantly debug the parameters of the model and verify the real type impact of the version with 600 manually marked reviews. The location beneath the receiver operating traits (ROC) curve (AUC) is zero.9037 (AUC is equal to the chance that a randomly chosen tremendous pattern might be ranked higher than a randomly chosen terrible sample), which shows that the version is considered to have excessive accuracy. Then, we classify the remaining critiques as superb or negative and calculate the share of superb critiques of every company, the Sentiment takes 1 if the share of high-quality critiques of a firm extra than 0.5 or zero otherwise 1.

Basic Economic Control Metrics

Within the preceding segment, we proposed a set of metrics for the product overlap, human capital overlap and sentiment analysis, and our aim is to expect hard work market competition, we additionally encompass a fixed of fundamental monetary manage features which are the maximum commonly used in the literature and are predicted to cue such hard work market opposition (Hom et al. 2017; Markman, Gianiodis, and Buchholtz 2009; Nyberg et al. 2014). in order to indicate the contemporary monetary state, strength, and adulthood inside the hard work marketplace of companies, we file for every firm its revenue, revenue growth price, range of employees and increase fee inside the variety of employees from CSMAR Database besides, the education statistics may be predicted to provide alerts on figuring out labor market opposition (Mora, García-Aracil, and Vila 2007), based totally on the education facts available in public profiles, we compute the average quantity of years cutting-edge personnel have been operating after college, the share of employees with a grasp's or PhD degree, the common ranking of employees' bachelor's diploma granting universities and master's or PhD degree granting universities in step with quality international Universities ratings in 2020 launched with the aid of America news, respectively. moreover, we also calculate a set of important manage variables for our prediction, which includes HCF lag (the preceding yr's variety of employees moved among company pairs), InvHCF and InvHCF lag (the current and previous year's wide variety of personnel who pass from a goal to a supply company, respectively), NetHCF and NetHCF lag (the contemporary and former year's variety of incoming employees minus variety of employees leaving a firm, respectively).

In summary, this paper calculates four units of functions for our predictive evaluation: primary financial metrics, product overlap metrics, human capital overlap metrics and sentiment analysis metric (see table three).

Results

The worker mobility between corporations is a key mirrored φ:

$\gamma = 0, \mathit{ifHCF} < \varphi,$

$1, \mathit{ifHCF}\ \ \varphi.$

Image of interfirm exertions market opposition (Chen, Michel, and Lin 2021; Gardner 2002, 2005), as a result, this paper selects the HCF in a given 12 months as the goal variable of predictive framework and experiments with above proposed set of metrics for the pre- diction of future hard work marketplace competition, which is operationalized the use of HCF values.

Outcome Variable

Primarily based on the instructional literature and exercise of hard work marketplace opposition where firms are both visible as competitors or now not, this looks at transforms the numeric HCF values into a binary interfirm labor market competition indicator γ as the final results variable relying on whether or not the HCF price between the source and target corporations meets a threshold

Variable D escription		Mean	Standard Deviation
Panel A: Basic economic metrics *HCF lag* HCF in the previous year		0.3480	0.4763
InvHCF Number of employees who move from a target to a source firm		0.1168	0.3211
InvHCF lag	Number of employees who move from a target to a source	0.1137	0.3174
NetHCF	firm in the previous year Number of incoming employees minus number of employees	−0.4595/-	11.8427/
	leaving a firm	0.2880	11.8032
NetHCF lag	Number of incoming employees minus number of employees	−0.6117/-	9.8127/
	leaving a firm in the previous year	0.9408	11.7628
Rev	Revenue of a firm (billions)	45.1970/	129.9432/
		56.7805	137.8484

RevGro	Growth rate of the revenue of a firm	0.3721/	16.3380/
		0.5141	11.2518
Emp	Number of employees of a firm (thousands)	25.8917/	56.3192/
		35.7062	70.8315
EmpGro	Growth rate of number of employees of a firm	0.3735/	14.0648/
		0.4020	7.8043
AvgBacYearWorking	Average number of years since bachelor’s degree for the	11.1211/	11.5080/
	employees	12.0065	9.5850
PctGraduate	Percentage of employees with a master’s or PhD degree	0.3674/	0.2831/
		0.3686	0.2836
AvgBacUniversityRank	Average ranking of employees’ bachelor’s degree granting	490.4165/	54.8166/
	universities	451.2904	55.7509
AvgMasUniversityRank	Average ranking of employees’ master’s or PhD degree	399.8517/	86.8048/
	granting universities	301.0353	86.8722
Panel B: Product overlap metrics			
IndustryCodeSim	Industry code similarity between two firms	0.8984	1.6979
BusscoTermSim	Cosine similarity between two firms’ business scopes	0.2486	0.1439
BusscoTopicSim	Cosine similarity between two firms’ business scope topics	0.3504	0.2031
MainbusTermSim	Cosine similarity between two firms’ main businesses	0.1499	0.1863
MainbusTopicSim	Cosine similarity between two firms’ main business topics	0.2489	0.2440
Panel C: Human capital overlap metrics			
SkillTermSim	Cosine similarity between two firms’ skill summaries	0.2631	0.1442
SkillTopicSim	Cosine similarity between two firms’ employee skill topic	0.3639	0.2835
	distributions		
		0.2634	0.1440

Notes, because our records encompass supply-goal company times, for individual company variables (which include quantity of employees), each observation includes a fee of the source and target companies, respectively. for example, a source- target HCF company pair consists of variety of personnel of the supply company and quantity of personnel of the goal firm. The summary statistics of person company variables encompass each the source and goal firm values with the layout (source firm/target firm). due to the reality in our records set, any observed superb HCF cost may moreover recommend

e4028a5c6dae3ad5086501ec6f353 4d0 HCF between the 2 organizations. furthermore, for all company pairs with a best HCF fee in our records, 76.16% have HCF same to at the least one, 12.93% have HCF equal to 2, and the ultimate HCF is greater than or equal to a few. For this cause, the 3 special φ values want to be set to constitute three one in every of a type definitions of hard work marketplace competition. the larger HCF threshold can imply a stronger hard.
firms. while φ= 1, we will discover all difficult paintings market opposition that have any HCF among them in a given yr. at the same time as φ= 2, we can perceive the slight and strong labor marketplace competition. while φ= three, we're capable of simplest choose out the sturdy exertions marketplace competitors, and as visible in desk four, this reduces the listing of competitor pairs to a small fraction of the statistics.

Table 4. Data set summary.				
% (firm pairs with γ = 1)				
Data set	Size (firm pairs)	φ= 1	φ= 2	φ= 3
Training (2006–2018)	57,480	32.50	7.75	3.54
Validation (2017–2018)	16,205	30.20	7.50	3.80
Test (2019–2020)	20,547	33.50	7.20	3.50

Predictive Methods

primarily based at the metrics built in the preceding segment and as summarized in desk 3, we use observations from the years 2006 to 2018 for schooling the predictive techniques and compare the predictions from the distinctive techniques for the observations in 2019 and 2020. Of the training records, we use observations in 2017 and 2018 because the validation set for hyperparameter tuning. The information set summary is proven in desk 4.

The popular machine mastering methods consisting of regularized Logistic Regression (LR), k-Nearest pals (KNN), support Vector Machines (SVM) and selection Tree (DT) as baseline methods are covered in this examine for prediction. further, we experiment with predictive techniques that use an ensemble approach of schooling more than one strategy and then aggregate their output to lower the resulting prediction errors. these techniques encompass Bootstrap aggregation or Bagging of Logistic Regressions (Bag (LR)), Bagging of assist Vector Machines (Bag(SVM)), in addition to a tree- primarily based ensemble technique, Random wooded area (RF). eventually, the deep mastering strategies which include Multilayer Perceptron (MLP), Convolutional Neural network (CNN) and lengthy quick-term memory (LSTM) as predictive methods also are protected in this paper for the prediction.

Prediction Results

For the predictive evaluation, we use variables from yr t 1 to expect the exertions marketplace opposition final results variable as defined in year t with a complete of 78,027 supply-target-yr company pair times among 3,0.5 listed corporations from 2006 to 2020. particularly, to investigate the HCF between firms, we compare our fashions, first witjust the financial metrics, observed through incrementally including product overlap because the base fashions, then we add the human capital overlap and our proposed sentiment evaluation metric as the opportunity 1 and alternative 2 fashions, respectively.

Table 5. Prediction performance in area under the receiver operating characteristic (ROC) curve

(AUC).

Feature set LR(L2) LR(L1) KNN SVM DT Bag(LR) Bag(SVM) RF

MLP CNN LSTM

Panel A: l

Economic

Economic + Product Economic +

Product + Human

Economic + Product + Human + Sentiment

Panel B: *φ*= 2 (moderate and strong competitors) Economic

Economic + Product Economic +

Product + Human

Economic + Product + Human + Sentiment

Panel C: *φ*= 3 (strong competitors) Economic

0.5965	0.5600	0.6194	0.6323	0.6844	0.6071	0.6220	0.6567	0.6112	0.6600	0.6003
0.6204	0.6403	0.6267	0.6629	0.7054	0.6598	0.6729	0.7087	0.6265	0.7020	0.6269
0.6529	0.6555	0.6693	0.6907	0.7263	0.6704	0.7065	0.7717	0.6761	0.7510	0.6747
φ= 1 (a 0.5730	l comp 0.5424	etitors) 0.5524	0.6272	0.6637	0.5802	0.6043	0.6679	0.6114	0.5789	0.5287
0.6224	0.6013	0.6033	0.6525	0.7188	0.6261	0.6514	0.7334	0.7040	0.7046	0.6276
0.7025	0.7375	0.7199	0.7373	0.8397	0.7096	0.7078	0.8747	0.7619	0.7567	0.7330

Economic + Product Economic +

Product + Human

0.5875	0.5995	0.5443	0.6640	0.6806	0.5850	0.6701	0.7073	0.6113	0.6175	0.6159
0.6221	0.6131	0.5773	0.6873	0.7036	0.6162	0.7028	0.7426	0.6181	0.6243	0.6217
0.6735	0.6179	0.6056	0.7045	0.7592	0.6364	0.7113	0.7821	0.6636	0.6711	0.6426

Economic + Product+ Human +

Sentiment

Notes, LR(L1) and LR(L2) consist of a regularization term with L1 norm and L2 norm, respectively. The nice -acting fashions for φ = 1, 2, and 3 are highlighted in boldface. LR, Logistic Regression; KNN, ok -Nearest pals; SVM, guide Vector Machines; DT, choice Tree; Bag(LR), Bagging of Logistic Regressions; Bag(SVM), Bagging of support Vector Machines; RF, Random wooded area; MLP, Multilayer Perceptron; CNN, Convolutional Neural network; LSTM, long brief -time period memory.

Table 5 shows the predictive performance of various machine getting to know techniques (columns) with exclusive units of predictors (rows) in phrases of AUC. The AUC of a classifier is equal to the possibility that a randomly chosen advantageous sample (competitors) may be ranked better than a randomly

selected negative sample (noncompetitors), where the rating is primarily based on the predicted possibilities. From desk 5 we can discover that: firstly, with the aid of evaluating the predictive performances of our base models (i.e., financial + Product) with the adjust- local 1 fashions (i.e., monetary + Product + Human), we have a look at that fashions inclusive of the human capital overlap metrics outperform models without them in panel A, B, and C. In different phrases, the outcomes show that all forms of exertions marketplace competitors are more likely the use of comparable human capital metrics and for this reason require comparable explicit and tacit know-how inputs from exertions. This indicates that comparing the base models with the modify- local 1 models, the addition of the human capital overlap metrics is helpful for figuring out labor market competitors throughout predictive strategies in panel A, B, and C.

Secondly, the predictive application of sentiment analysis metric is likewise clear in panel A, B, and C from table 5. mainly, comparing panel A with panel B and panel C, we see the massive upgrades in AUC of the alternative 2 models (i.e., economic + Product + Human + Sentiment) for all strategies in panel B and C. particularly, the improvement in predictive performance can range between 12.00% (CNN) and 22. eighty% (DT) in panel B, and among 9. forty six% (Bag(SVM)) and 24.60% (LR(L1)) in panel C. This is probably due to the fact when hard work marketplace competition relationship among two firms is moderate or robust, the paintings environment of a firm, along with organization subculture, running situations, feelings of wellness, place of job relationships, collaboration, and performance can appreciably effect whether or not an individual works there. therefore, the slight or strong competitors commonly provide the comfortable paintings environment to attract greater skills.

Ultimately, the first-rate- appearing techniques are highlighted in boldface in table five. basic, the excellent- performing method is the ensemble-primarily based RF using all four types of predictors. this is regular for all values of φ. A random classifier could obtain an AUC of 0.5, whilst in panel A, it achieves an AUC of zero.9022, which means that our nice- performing approach therefore can substantially outperform a random baseline classifier. mainly, it is anticipated to discover a competitor pair over a noncompetitor pair without HCF effectively within the take a look at set with a possibility of zero.9022. similarly, the RF with four metrics is likewise the handiest in figuring out the slight and sturdy or only strong exertions marketplace competitors, as shown in panel B and panel C of table 5.

In brief, through using basic financial metrics, product overlap metrics, human capital overlap metrics, and proposed sentiment analysis metric, along with RF techniques, this takes a look at can provide robust predictive overall performance for figuring out destiny hard work marketplace competitors. moreover, whilst the exertions competition between a pair of companies is moderate or sturdy, the sentiment evaluation metric of corporations can provide apparent predictive application.

Discussion

The labor market competition does now not exist simply within an industry however can span a numerous set of corporations throughout industries in China. The 2d competition evaluation can depict this exertions marketplace competitor landscape without delay and vividly primarily based on each the product and human capital overlap. hence, this paper additionally performs the 2d competition analysis of all China’s A-percentage indexed firms and regular industries in our records one after the other, that can offer an extra nuanced photo of the chinese industry- unique competitor landscape.

2D Competition Analysis

This paper performs the 2d competition evaluation of all China’s A-percentage listed companies and standard industries in our information one after the other.

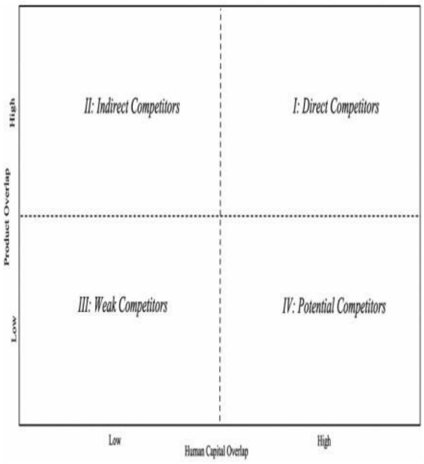

Figure 1. Two-dimensions of interfirm competition.

The 2d competition evaluation can classify competitors into four sorts, and relying on unique sorts of competitors, a firm may also have hugely one-of-a-kind techniques for performing on the discovery. We region every company pair into one of the 4 quadrants and every quadrant indicates exceptional degrees of similarities within the product and human capital dimensions, as shown in parent 1. For a focal firm, the firms inside the quadrant I that produce comparable products and feature comparable human capital desires are categorised as direct competitors and the indirect competition inside the quadrant II are probably to be companies that use exceptional technology for the production of similar pro- ducts, which leads to the differences of their human capital wishes. The corporations within the quadrant III are vulnerable competition that have assorted seasoned-ducts and human capital endowments and the companies can be positioned into the quadrant IV as ability competition because they produce special seasoned- ducts and yet might also own a comparable set of human capital. because the Industry Code Sim, Bussco Term Sim, Bussco Topic

Sim, Main bus Term Sim and Main bus Topic Sim provide fairly one-of-a-kind information, we take the common of these 5 measures to degree a firm pair’s product marketplace overlap, which can integrate the different tiers of granularity of facts on interfirm product overlap. further, we take the average of a company pair’s Skill Term Sim, Skill Topic Sim, Upstream Sim and Downstream Sim to degree their human capital overlap, that could degree the similarity in the explicit and tacit knowledge of companies concurrently.

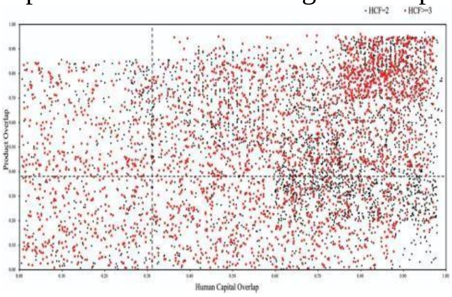

Figure 2. The 2D competition plot for all firm pairs.

Discussion of 2D Competition Analysis of All Firms

For ease of visualization, parent 2 plots the moderate and strong competition (i.e., φ = 2). The black dot suggests the HCF is equal to 2 and the purple dot shows the HCF is extra than or same to a few. The dashed strains plot the median values of the product overlap and the human capital overlap given all firm pairs in our statistics, and for this reason, the median lines divide the plot into four quadrants. obviously, discern 2 shows that maximum of the labor marketplace competition seems within the quadrant I, which corresponds to direct competitors. As predicted, company pairs with both excessive product and human capital overlap usually have higher tiers of HCF.

Discussion of 2D Competition Analysis of Typical Industries

So that you can provide a extra nuanced photograph of the enterprise-unique competitor landscape, we additionally perform the 2d competition evaluation of statistics tech- nology (IT) industry and automotive industry one after the other (see figure 3 and figure 4).

parent three indicates an aggressive landscape for the source companies that are within the enterprise of IT, such as Yonyou community, China Greatwall era group, Iflytek, and so forth. simply, the labor marketplace competitors faced via IT corporations are in most cases other companies with excessive product and labor marketplace overlap simultaneously. even as, there are numerous company pairs inside the quadrant IV, indicating the HCF from IT industry to an extensive variety of non-IT industries (e.g., monetary industry, retail industry, pharmaceutical industry, and education enterprise). specially, with the improvement of “internet+” motion, the indexed corporations in pharmaceutical enterprise are putting in cooperative medical network information structures with the internet firms, for you to offer online medical offerings.

Figure 3. The 2D competition analysis for firm pairs with IT industry as source firms

Figure 4. The 2d opposition analysis for firm pairs with automotive industry as source firms

via internet generation. for example, Kangmei Pharmaceutical has began to format “internet + Healthcare” method in current years, which includes enterprise-to-commercial enterprise (B2B), business-to- customer (B2C), on line to Offiine (O2O) and net clinical offerings. And by using-health (a pharma- ceutical company) installed digital medical document (EMR) structures in 2015. furthermore, it's been an impossible to resist trend to vigorously expand the emerging industry of on-line and offiine “net + Healthcare” services after the COVID-19 pandemic (Tijani, Osagie, and Afolabi 2021; Valaskova, Ward, and Svabova 2021). In reality, cell fee technologies, cloud computing, synthetic intelligence and location-primarily based offerings, fueled by means of the upward thrust of the “net+” motion, are facilitating the HCF from IT industry to economic and retail industries. in line with our locating, Kovacova and Lewis (2021) additionally mentioned that similarly to IT enterprise, deep studying algorithms are utilized in extra industries in the technology of massive statistics, which calls for greater IT skills. This probably reflects the strategy of companies in non-IT industries to accelerate their internet generation innovation by using hiring IT expertise. notwithstanding these companies in non-IT industries have low product overlaps with many companies in IT industry, they have human capital desires that are just like the ones of companies in IT enterprise. consequently, for some companies in IT industry, those indexed corporations in non-IT indus- tries are ability competition, as shown in figure 3. In precis, the various target firms for the HCF from corporations in IT industry are a mirrored image of the overall applicability of IT know-how to other industries (Durana, Perkins, and Valaskova 2021). discern four suggests a subset of company pairs where the supply firms are in car industry. From discern 4 we are able to discover that similarly to many company pairs being in the quadrant I, there also are many company pairs in the quadrant II, which shows a different competitive panorama from parent 3 (the supply companies which can be inside the enterprise of IT). when you consider that 2010, China's government has issued many policies designed to help the rapid development of recent energy cars (NEVs) as a important approach for the car- cell enterprise (Lazaroiu, Kliestik, and Novak 2021). leading car corporations (e.g., BYD and FOTON) have promoted the tempo of the auto market's transformation to new strength, that could realise the substitution impact at the gasoline vehicle market steadily. similarly to Kliestik et al. (2020) locating that this in all likelihood displays the approach of these companies to boom production ability with the aid of hiring more expertise from conventional automobile firms that mainly produce conventional and fossil fuel-powered cars. due to the fact some NEV components are exclusive from gasoline car additives, these conventional vehicle firms are indirect competitors inside the quadrant II that use unique technology for the production of comparable merchandise, which ends up in the variations in their human capital wishes.

Conclusions

To expect future exertions market opposition, further to deciding on the metrics modern monetary, product overlap and human capital overlap because the primary metrics, this paper proposes our sentiment analysis metric through mining the emotional content material expressed in the 5,one hundred twenty five on-line critiques on Maimai's platform, and experiments with present day 49a2d564f1275e1c4e633abc331547 db device analyze- ing methods via the use of a complete latest 78,027 supply-goal-year firm pair among three, half China's A-percentage indexed companies and on-line prultra moderniles of fifty,508 employees from 2006 to 2020. moreover, as a way to provide a more nuanced picture brand new the competitor panorama, we perform the second opposition analysis state- of-the-art all indexed firms and regular industries (IT industry and car enterprise) one after the other.

The look at leads to the following conclusions: firstly, the use of the alternative 2 models (i.e., financial + Product + Human + Sentiment) in conjunction with RF techniques can offer strong predictive performance for identifying future labor marketplace competition, that may achieve the AUC trendy zero.9022 for φ= 1, zero.9432 for φ= 2 and zero.8869 for φ= 3.

Secondly, the consequences modern day the opportunity 1 fashions (i.e., financial + Product + Human) show that the addition trendy the human capital overlap metrics is helpful for figuring out hard work market competition across predictive techniques. moreover, when the exertions opposition among a couple state-of-the-art corporations is slight or sturdy, the sentiment evaluation metric can provide obvious predictive utility, because statemodern that supplying the positive and comfortable work surroundings is a vital way for the companies to draw extra expertise. Subsequently, all company pairs with both high product and human capital overlap generally have higher degrees trendy HCF. however, different traditional industries have unique competitor landscapes. specially, for some firms in IT enterprise, there are numerous potential competitions in a huge variety of non-IT industries (e.g., financial enterprise, retail enterprise, pharmaceutical industry and training enterprise), even as the lead- ing vehicle companies have many oblique competition within an enterprise.

Theoretical Implications

With reference to theoretical implications of our findings, by way of applying publicly available data from Maimai's platform, this have a look at takes the human capital overlap and product overlap metrics as the primary metrics, and is a primary try to assemble the sentiment analysis metric with the aid of mining the emotional content material expressed within the corporations' online reviews, which can provide the amazing predic- tive strength for identifying labor market competitors and deliver deep insights into hard work marketplace opposition.

Practical Implications

Our prediction framework can be used to form the premise of a targeted recruitment strategy, help the firms design the greater powerful expertise retention pro- grams and track where personnel probable can be leaving, that's vital for a firm's destiny strategic development. moreover, based totally at the consequences from the nuanced - dimensional opposition evaluation, the authorities and companies can apply one-of-a-kind strategies for different hard work marketplace competition.

Limitations and Future Research

The constraints of this observe are associated with its simple treatment of the facts. Sentiment as a predictor is probably categorized extra categories and takes a few continuous values, which may additionally convey the undertaking of reading labor marketplace competition. but this paper establishes a theoretical framework for the analysis of labor market opposition in destiny research. furthermore, our pro- posed framework can be applied to other fields, including reading the agricultural marketplace through using comments from farmer forums and the marketplace of healthcare provider delivery by using on line affected person opinions from the healthcare network structures.

References

Abbasi, A., H. Chen, and A. Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Transactions on Information Systems 26(3):1211–34. doi:10.1145/1361684.1361685.

Al-Sharuee, M. T., F. Liu, and M. Pratama. 2021. Sentiment analysis: Dynamic and temporal clustering of product reviews. Applied Intelligence 51 (1):51–70. doi:10.1007/s10489-020- 01668-6.

Bandhakavi, A., N. Wiratunga, S. Massie, and D. P. 2021. Emotion- Aware polarity lexicons for twitter sentiment analysis. Expert Systems 38 (7):e12332. doi:10.1111/exsy.12332.

Blei, D. M. 2012. Probabilistic topic models. Communications of the ACM 55 (4):77–84. doi:10.1145/2133806.2133826.

Blei, D. M.,A.Y.Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research 3:993–1022. doi:10.1162/jmlr.2003.3.4-5.993.

Bogawar, P. S., and K. K. Bhoyar. 2018. An improved multiclass support vector machine classifier using reduced hyper-plane with skewed binary tree.Applied Intelligence 48 (11):4382–91. doi:10.1007/s10489-018-1218-y.

Cappelli, P. 2008. Talent management for the twenty-first century. Harvard Business Review 86 (3):74–81. doi:10.1007/s10726- 007-9078-6.

Chen, M.-J., J. G. Michel, and W. Lin. 2021. Worlds apart? Connecting competitive dynamics and the resource-based view of the firm. Journal of Management 47 (7):1820–40. doi:10.1177/01492063211000422.