

Exploring enterprises competition: From a perspective of massive recruitment texts mining

Lu Liu¹

Jiaming Fang^{2,*}

Liangqiang Li³

*Corresponding author

¹ PhD student, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, lliu@std.uestc.edu.cn

² Professor, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, jmfang@uestc.edu.cn

³ Lecturer, Sichuan Agricultural University, Dujiangyan, Sichuan, China, lilq@sicau.edu.cn

ABSTRACT

Extant research has made limited efforts to conduct competitive intelligence analysis based on recruitment texts. To fill the gap, this study proposes a method for deriving and analyzing competitive relationships, identifying competition paths, and calculating asymmetric competitiveness degrees, from the recruitment texts on e-recruiting websites. Specifically, this study developed a competitive evaluation index system for companies' skill needs and resource base based on 53,171 job descriptions and 42,641 company profiles published by companies across 8 industries (including 35 industry segments) using automated text processing methods. Furthermore, in order to identify competitive paths and calculate the degree of asymmetric competitiveness, this study proposes a modified bipartite graph approach (i.e., MBGA) for competitive intelligence analysis of recruitment texts based on the competition evaluation index system. Experiments on a real-world dataset of the representative companies clearly validated the effectiveness of the method. Compared to the five state-of-the-art methods, MBGA performs better in disclosing the overall competition and is more accurate in terms of the error rating ratio (i.e., ERR) of the competition.

Keywords: Recruitment texts, competitive intelligence, asymmetric competitiveness degrees, bipartite graph.

INTRODUCTION

As the number of enterprise registrations continues to grow and the boundary of industries becomes increasingly blurred, the enterprises and even cross-industry competition have become intense and diverse. With the emergence of new technologies and the development of resources endowment, the overlap in skills demand and resource supply among enterprises has become ubiquitous, as well as competitors are everywhere. In order to adapt to the change market, enterprises must quickly and accurately find rivals, and then clear competitive strategy, own competitive (dis)advantages and specific competitive means (Jantunen *et al.*, 2005).

With the development of Internet, enterprises not only have a broader gathering place, but also leave many traces and information. Also, there is some information with competitive analysis attributes, for example, products and performance information, user search logs and online reviews (Netzer *et al.*, 2012; Ringel & Skiera, 2016; Wei *et al.*, 2016). Based on market commonality of skills demand and resource similarity of resources supply, job descriptions and enterprise profiles contained in recruitment texts can reflect enterprises competition. However, it is obviously not possible to deal with massive recruitment texts manually. Thus, the intelligent text processing method is needed to competitive intelligent analysis between enterprises. Additionally, based on market commonality and resource similarity theories, this study proposes the skills demand for talents market and resources supply about enterprises endowment also have the competitive analysis attributes, which has not been adequately explored in competitive intelligence field.

Based on the above background and extant studies, the following problems still need to be solved. First, how to use intelligent methods to mine competitive information in massive recruitment texts? Second, how to use the competitive information mined above to analyze enterprises competition? In order to solve these problems, this paper uses massive recruitment texts published by enterprises to extract skills demand and resources supply indicators by using text processing methods of Word2Vec and K-means. Then, this paper uses the above indicators and bipartite graph method to visualize competition relations, identify competition paths, and calculate asymmetric competitiveness degrees. Finally, we use enterprises real data to test the effectiveness and advantages of above methods, as well as put forward support for competition strategy.

LITERATURE REVIEW

Research on Enterprises Competition

The research on enterprises competition can be divided into two categories, one is to identify and analyze enterprises competition from marketing and management perspective (Long *et al.*, 2010), the other is from data mining perspective, which proposes how to mine competition information from big data (Gutt *et al.*, 2019; Zhong *et al.*, 2018). This research belongs to the latter, which is to find implicit competition information and knowledge through massive recruitment texts mining, and then to provide support for competition decision. Accordingly, this part review includes competition analysis theories and methods.

The competition analysis theories involve three aspects: market demand commonality or overlapping (Xu *et al.*, 2018), resources supply similarity (Barney, 1991), and integration of demand and supply (Ma, 2015). This study belongs to integration theory perspective. Specifically, this study combines market commonality of skills demand and resource similarity of resources supply. Chen (1996) put forward the theories of market commonality and resource similarity in enterprises competition. The market commonality can be defined as the degree of overlap between enterprises in market, and the definition is mostly based on products or consumers (Chen, 1996). In addition, market is a complex and multidimensional construct that can be adapted to different strategic purposes (Porac & Thomas, 1990). Accordingly, the overlaps demand of talents or skills market also form competition relations. Resource similarity can be defined as types and quantities comparison of strategic resource endowments between enterprises, or the degree of competitors' resource endowments that can compete with focus enterprises (Deng *et al.*, 2017). Enterprises with similar resources may compete for similar strategic capabilities (Chen, 1996). Besides, asymmetric competition is also the focus of this study, that is, threat and competitiveness degree of A to B may be different from B to A (Ringel & Skiera, 2016). In other words, an enterprise can be exclusive in some way, and it cannot be easily replaced by other. This can be interpreted as: there is another exactly the same enterprise, and the exclusivity of focus enterprise is indicated by competitiveness degree between these two identical enterprises. The specific concepts and definitions are shown in Table 1.

Table 1: Definitions of relevant concepts.

Concepts	Author(s) (Year)	Definition / Description
Market Commonality	Chen (1996)	"The degree of presence that a competitor manifests in the markets it overlaps with the focal firm."
Resource Similarity	Chen (1996)	"The extent to which a given competitor possesses strategic endowments comparable, in terms of both type and amount, to those of the focal firm."
Asymmetric Competition	Ringel and Skiera (2016)	"Competitive asymmetry exists when the degree of competition between two firms is not equal, as when Firm A competes more intensely with Firm B than Firm B competes with Firm A."

The competition analysis methods can be divided into relatively traditional and intelligent methods. The former is more focused on theory construction and model building (Peng & Liang, 2016), while the latter (i.e., competitive intelligence) is more focused on extracting competitive content and focused on method innovation (Ringel & Skiera, 2016). For example, Dai *et al.* (2018) used user coincidence to identify competitors based on consumer behavior data, and proposed a competitor identification method that can adapt to dynamic market. However, this method cannot be used in low frequency consumption scenarios. Pant and Sheng (2015) predicted competition relations based on in-links and out-links of websites. Furthermore, studies identified competition relations by building and measuring network structures (Gupta *et al.*, 2019; Ma *et al.*, 2011).

In summary, the competition analysis theories can be summed up as market commonality and resource similarity. Existing studies are based on one of them, and there are few studies on competition analysis from the integration theory. The competition analysis methods can be divided into traditional and intelligent methods, such as case analysis and text mining. Traditional methods are being challenged by lack of dynamic, and intelligent methods have gradually become competition research focus. Therefore, this research attempts to integrate the two theories, and uses intelligent text processing methods for enterprises competition analysis.

Research on Application of Recruitment Information

Recruitment information is an audience-specific message that usually includes position, job responsibilities, qualifications, and enterprise profile, etc. It not only presents the demand of talents and skills, but also the supply of resources, products, and services. So, human resources demand, talents qualification, technologies application and resources endowment can be got by text mining, and even strategic development information can be found. Moreover, massive recruitment texts can be used for competition analysis. This can be reviewed in two ways, that is, extant application contexts of recruitment information mining and competition analysis attributes of various massive texts. Accordingly, this part review includes application contexts of recruitment information mining and competition analysis attributes of different massive texts.

It has become research direction of many scholars to mine recruitment texts, and then to discover potential knowledge as well as to apply it in different contexts. For example, by mining recruitment information, it can identify multifaceted nature of job skills (Gardiner *et al.*, 2018), derive job trends (Smith & Ali, 2014), and adopt association rules to predict industry (Zhang *et al.*, 2017), and so on.

In addition, some studies explored competition analysis attributes of different data (Mohammed *et al.*, 2014). For example, Wei *et al.* (2016) used same functional requirements of different brands in query logs to measure competitiveness degree among enterprises and brands. Qiao *et al.* (2017) mined keywords relationship from user query logs, and then identified competitive keywords to help enterprises make advertising competitive strategy in search engine.

In summary, studies on recruitment information application and competition analysis attributes are shown in Table 2. The former provides a text processing basis for this research. However, there is a lack of semantic understanding in the text processing methods. Additionally, the research of competition analysis attributes about massive texts provides a logical basis for this study. The recruitment information is the same as other kinds of information used for competition analysis, and has competition analysis attributes. However, most of the mentioned above studies focus on jobs, skills analysis, and not explore enterprise level relations. Therefore, this research attempts to mine and apply recruitment texts to analyze enterprises competition.

Table 2: Studies on recruitment information application and competition analysis attributes.

Author(s) (Year)	Application of Recruitment Mining	Contexts of Information	Author(s) (Year)	Competition Attributes of Different Texts	Analysis of Massive Texts
Smith and Ali (2014)	Job trends		Netzer <i>et al.</i> (2012)	UGC	
Zhang <i>et al.</i> , (2017)	Industries association predict		Chakraborti and Dey (2014)	Multi-document Text	
Gardiner <i>et al.</i> (2018)	Multifaceted nature of skills		Ringel and Skiera (2016)	Query logs	
Turrell <i>et al.</i> (2019)	Jobs classification		Nam <i>et al.</i> (2017)	Social tags	
Wu <i>et al.</i> (2019)	State of enterprises aggregation		Gao <i>et al.</i> (2018)	Online reviews	

RESEARCH METHODOLOGY

This study aims to analyze competitive intelligence through recruitment texts mining with utilization of several methods together. The research logic used for this study is illustrated in Figure 1. First, we performed text mining methods using Word2Vec and K-means which is widely applied in the existing text similarity studies. Competitive content was extracted from job descriptions and company profiles, and the evaluation index systems of skills demand and resource base were constructed. Furthermore, we applied a modified bipartite diagram approach to create a bipartite diagram of firm-skills and firm-resources based on the indicators in the indicator system. Then, competitive paths of skills and resources among companies are identified. Based on the competitive path and competitive intensity, the asymmetric competitive degrees between firms are calculated and the asymmetric competitiveness degrees matrix is obtained.

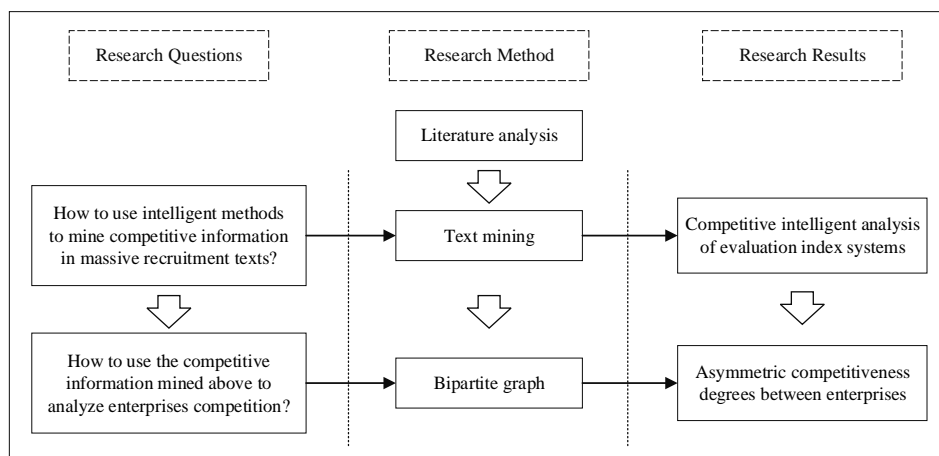


Figure 1: The logic framework of this research.

Data Source

This study used a crawler written in python to collect 53,171 job descriptions and 42,641 company profiles from a well-known recruitment platform called Zhaopin.com from December 2018 to January 2022. The descriptive statistics of companies are shown in Table 3. Founded in 1994, Zhaopin.com has evolved into the leading career development platform in China, providing relevant career and development opportunities for users throughout their careers. The crawling web information is shown in Figure 2. The original texts of "Job Description" and "Company Profile" are in Chinese and have been translated into English as shown in Figure 2.

Table 3: Descriptive statistics of companies.

Category	Subdivision	No.	Accumulative No.	Accumulative %
Industry	IT	6,887	6,887	16.15
	Finance	2,223	9,110	21.36

	Real estate and construction, manufacturing	13,480	22,590	52.98
	Retailing and logistics	2,412	25,002	58.63
	Education media and advertising	5,817	30,819	72.28
	Services	5,875	36,694	86.05
	Marketing and sales	2,806	39,500	92.63
	Personnel administration	1,285	40,785	95.65
	Others (e.g., farming, forestry, husbandry and fishing, etc.)	1,856	42,641	100
Type	Private	28,145	28,145	66.00
	Joint venture	4,753	32,898	77.15
	State-owned	2,286	35,184	82.51
	The listed	1,985	37,169	87.17
	Joint	1,504	38,673	90.69
	Foreign-invested	1,034	39,707	93.12
	Others (e.g., School, Hospital, Bank, etc.)	2,934	42,641	100
Established	2012-2021	9,241	9,241	21.67
	2002-2011	8,428	17,669	41.44
	1992-2001	4,583	22,252	52.18
	1982-1991	1,047	23,299	54.64
	Others (e.g., other years, and not shown)	19,342	42,641	100
Employees	≥ 10000	1840	1840	4.32
	1000-9999	4701	6541	15.34
	500-999	3645	10186	23.89
	100-499	13,207	23,393	54.86
	≤ 99	18,100	41,493	97.31
	Not shown	1,148	42,641	100
Jobs	≤ 5	12,248	12,248	28.72
	6-10	26,970	39,218	91.97
	11-20	868	40,086	94.01
	21-50	859	40,945	96.02
	≥ 51	1696	42,641	100

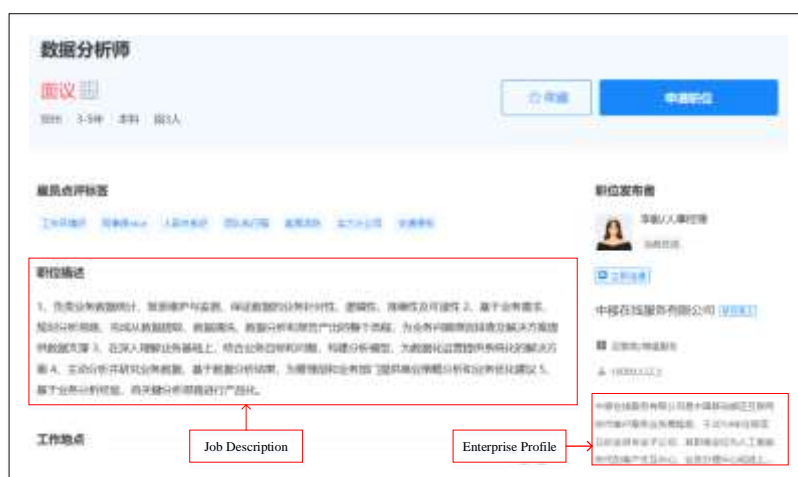


Figure 2: Example of crawling web information.

Text Processing

The purpose of this part is to preprocess and cluster the above recruitment texts. First, Jieba (i.e., Chinese text segmentation tool) with industries lexicons is used to clean data. Specifically, 58 professional industries lexicons about Internet, finance, real estate, trade, retail, education, media, advertising, logistics, marketing, personnel and administration, and so on, from Sogou dictionary, which is often used to improve text segmentation effect (Yan & Ma, 2019), were obtained (An *et al.*, 2019). Next, text mining methods are used to make a meaningful division of above words. It is divided into three steps: (1) using Word2Vec to train word vectors; (2) using K-means to calculate distance between the vectors to divide the words; (3) getting cluster results of skills and resources. The method framework of competitive information mining is shown in Figure 3.

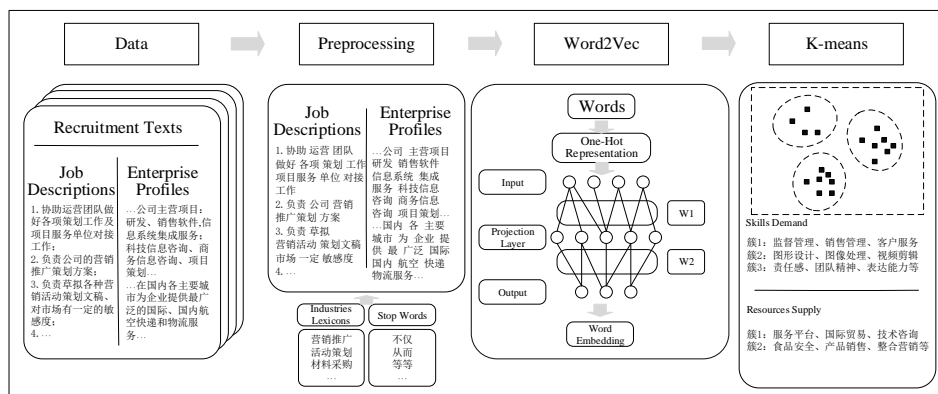


Figure 3: Methods framework of competitive information mining.

CONTENTS AND RESULTS

This part mainly includes three aspects: (1) constructing the competition evaluation index systems of skills and resources; (2) using the index systems combined with bipartite graph to visualize competition relations, identify competitive paths, and calculate asymmetric competitiveness degree between enterprises; (3) using enterprises real data to test validity and advantages of above method, and to provide support for enterprise competition strategy and evaluation.

Construction of Competition Evaluation Index Systems about Skills and Resources

The purpose of this part is to construct competition evaluation index systems of skills demand and resources supply based on cluster results of the above text processing.

Competition evaluation index system of skills demand

To further demonstrate cluster results, this study constructs an index system according to industry classification of zhaopin.com. It is divided into IT, finance, real estate and construction, retail and logistics, education media and advertising, services, marketing and sales, personnel administration. In addition, the industry classification of the platform is instructive, scientific and representative, and the industry classification of other platforms is similar to it. Then, cluster results are mapped to technical/core skills, business/auxiliary skills, and quality/comprehensive skills indicators to show focus and generalization of different skills, and to increase intuitiveness and practicality. The examples of representative indicators of skills demand index system are shown in Table 4.

Table 4: Examples of representative indicators of skills demand index system.

Industry	Technical/Core Skills	Business/Auxiliary Skills	Quality/Comprehensive Skills
IT	Data Analysis	Security control, machine learning, programming technology, programming language, parallel computing, program development, storage technology, etc.; Baidu bidding, shopping search, keyword search, bidding ranking, media planning, media promotion, brand planning, brand marketing, etc.;	security awareness, reasonable suggestion, technical support, information feedback, strict implementation, remedial measures,
	New Media Operation		
Finance	Financial Research	Financial investment, risk management, financial statements, accounting calculation, stock trading, equity investment, property management, etc.;	honest, dependable, trustworthy, take risks, abstract thinking, innovation ability, independent thinking,
	Investment Analysis		
Real Estate and Construction	Architectural Design	Quality control, urban planning, engineering design, process inspection, architectural planning, architectural energy conservation, architectural evaluation, etc.;	dedication, communication, mutual assistance, cooperation, active participation,
	Real Estate Planning		
Retail and Logistics	PM	Material procurement, warehouse receipt, regional agent, price comparison, capital input, capital support, inventory reduction, etc.;	lubrication, quick reaction, logical thinking, collaboration,
	Logistics Operations		
Education Media and	Post Production	Advertising design, advertising agency, technical consultation, content analysis, content production, image processing, graphic	fighting spirit, responsible, adaptability, good attitude,

Advertising	Visual Design	processing, image design, etc.; Glass-fronted billboard recommendation, activity promotion, structure design, interface design, landscape design, space design, art design, etc.;	team spirit, strategic thinking, self-management, rigorousness, etc.
Services	Travel Advice	Consulting service, regional analysis, market information, multidimensional analysis, service consultation, improvement service, customer service, etc.;	
	Hotel Management	Industry analysis, resource management, organization setup, decision analysis, decision support system, sustainable development, practicability analysis, etc.;	
Marketing and Sales	Marketing	Procurement planning, product development, market research, market promotion, pricing strategy, pricing model, advertising planning, etc.;	
	Internet Marketing	Viral marketing, interactive marketing, campaign promotion, WOM marketing, social marketing, online sales, etc.;	
Personnel Administration	HRM	Performance appraisal, performance evaluation, performance assessment, skills training, performance measurement, personnel evaluation, recruitment, etc.;	
	Legal Consultation	Mergers and acquisitions, property security, legal services, annual plans, tax policies, information research, online consulting, strategic planning, etc.	

Tips: PM means Procurement Management; HRM means Human Resources Management.

Competition evaluation index system of resources supply

According to resources endowment and resources utilization capacity in Resource-Based View, this paper demonstrates cluster results. According to characteristics of enterprise resources: value, rareness, imperfect imitability and substitutability (Barney, 1991), and viewpoint that resources are input of production process and capability is efficiency of resources utilization, this study designs initial classification criteria of indicators. According to viewpoint that only by integrating resources purposefully can resources become the source of competitive advantage, and extant core competence of enterprises, namely business environment, enterprise, subject, technology, product and core product subsystem (Wang *et al.*, 2000), the study designs ultimate classification criteria of indicators. The examples of representative indicators of resources supply index system are shown in Table 5.

Table 5: Examples of representative indicators of resources supply index system.

Resources Endowment	Indicators	Enterprise Capacities	Indicators
Technology Resources	Office automation, financial management, test center, product inspection, product design, project contracting, program development, financial accounting, financial consulting, color printing and packaging, multimedia technology, molecular biology, structural design, molding technology, software testing, business intelligence, QIC, MES, professional translation, integrated marketing, information security, identification system, biological prevention and control, digital technology, hydraulic engineering, network resources, etc.;	Resource Utilization Capacities	Exclusive agency, sole proprietorship, leading position, domain expert, leading enterprise, authoritative authentication, national top ten, the listed company, provincial key, world leading, global strategy, international finance, state-owned bank, state-owned capital, joint ventures/enterprises, leading brands, authoritative departments, authoritative institutions, trinity/three-in-one, joint ventures, foreign sole ownership, Chinese-western combination, directly affiliated department, environment friendly, industry pilot, five-in-one, pillar industries, well-known brands, well-known trademarks, general distributors, well-known enterprises, etc.;
Natural Endowments	Red/Revolutionary tourism, mineral resources, national characteristics, beautiful scenery, ecological landscape, scarce resources, renewable energy, renewable resources, Chinese culture, sightseeing agriculture, central location, etc.;		Safety equipment, safety glass, packaging materials, glass fiber, heating system, TCFL, synthetic rubber, care products, integrated circuits, hotel supplies, adhesive products, air conditioning equipment, access control systems, automotive parts, green building materials, semiconductor device, insulation materials, glass curtain wall,
Industry Positioning	Health food, insurance industry, tutorial school, editing and publishing, hotel, financial consulting, catering industry, coal industry, food industry, municipal engineering, film and television production, examination-oriented education, housing industry, online education, early education, information industry, medical cosmetology, pharmaceutical industry, preschool, garden engineering, legal consulting, higher education,	Core Products	

	outdoor products, household products, etc.;		financial software, diesel engines, electrical equipment, elevator accessories, protective supplies, non-metallic materials, clothing, composite materials, mobile phone accessories, etc.
Market Area	Large and medium cities, large, medium and small cities, three northeast provinces, Southeast Asia, whole country, Shenzhen sez, Hong Kong and Macao regions, Hong Kong, Macao and Taiwan regions, Guangdong region, Chengdu Region, Europe and America region, Zhangjiajie, Pearl River Delta region, Tibet region, etc.;		

Tips: QIC means Quality Inspection Center; MES means Manufacturing Execution System; TCFL means Teaching Chinese as Foreign Language.

Competition Analysis between Enterprises Based on Bipartite Graph

The purpose of this part is to visualize competition relations, identify competitive paths, and then to quantify asymmetric competitiveness degrees between enterprises, using indicators of skills demand and resources supply in the above index system, as well as bipartite graph.

Bipartite graph method has been widely used in relation recognition between entities (Choi *et al.*, 2022; Liu *et al.*, 2020; Wei *et al.*, 2016). Inspired by Wei *et al.* (2016) research logic, this research following the logic expands and verifies bipartite graph method in competitive intelligence field. In addition, this research made the following improvements to the method: First, this study uses multisource texts (i.e., job descriptions and enterprise profiles from recruitment texts) based on market commonality and resource similarity theories, to verify the validity of the method. Second, this study adjusts the competition paths identification and asymmetric competitiveness degrees calculation based on the multisource texts, to expand the applicability of the method.

Bipartite graphs construction of competition relations

This study uses bipartite graph to analyze competition, that is, visualize competition relations between enterprises in skills demand and resources supply. Firstly, two bipartite graphs of enterprises-skills and enterprises-resources are constructed by using nodes with different attributes and edges between the nodes. Then, based on the above bipartite graphs and the evaluation index systems, the frequency, which represents enterprise's demand for the skill or supply for the resource mined from job descriptions and enterprise profiles respectively, of each edge can be indicated. It can be defined as follows:

Definition 1: Given three sets of entities $E = \{e_1, e_2, e_3, \dots\}$, $S = \{s_1, s_2, s_3, \dots\}$, and $R = \{r_1, r_2, r_3, \dots\}$ respectively representing enterprises, skills and resources, the entities with different attributes can be connected by L_{ES} and L_{ER} , and then bipartite graphs of enterprises-skills $B_s = \{E, S, L_{ES}\}$ and enterprises-resources $B_r = \{E, R, L_{ER}\}$ can be build.

Where E is enterprises set, S is skills demand set corresponding job descriptions, and R is resources supply set corresponding enterprise profiles. L_{ES} is a set of edges, defined by the focus enterprise corresponding skills demand and frequency, and L_{ER} defined by the focus enterprise corresponding resources supply and frequency. Structurally, E , S , and R represent three nodes set, i.e. enterprises, skills, and resources, as well as L_{ES} and L_{ER} represent two edges set, i.e. E and S or E and R . Next, this research uses above definition and overall enterprises information mined from job descriptions and enterprise profiles, as shown in Table 6, to construct enterprises- skills and enterprises-resources bipartite graphs, as shown in Figure 4.

Table 6: Bipartite graphs construct information (an illustrative example).

Enterprises	Skills Demand	Frequency	Resources Supply	Frequency
e_1	Product marketing	12	Information system	9
	Data analysis	10	Clients database	8
	Website OM	9	Marketing tactics	8
e_2	Product marketing	15	Information system	9
	Video editing	11	Marketing tactics	12
	Website OM	10	Patent technology	11
e_3	Data analysis	13	Clients database	11
	Video editing	12	Patent technology	13
	Webpage production	15	International trade	10

Tips: Website OM means Website operation and maintenance.

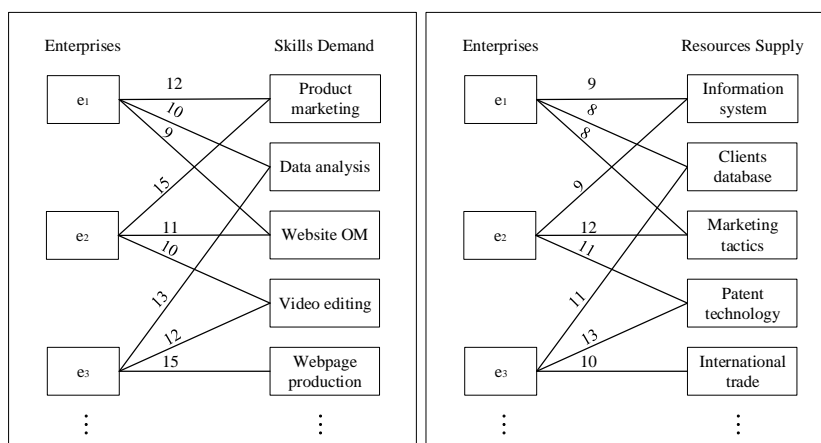


Figure 4: Bipartite graphs of skills demand and resources supply (an illustrative example).

Competition paths identification based on the above bipartite graphs

In order to measure asymmetric competitiveness degrees between enterprises, after constructing the bipartite graphs, we must identify competition paths, i.e., overlap in which skills are demanded or resources are supplied, between paired enterprises. In other words, connecting two enterprises with the same skills demand or resources supply, it can be defined as follows:

Definition 2: Given $B_S = \{E, S, L_{ES}\}$ and $B_R = \{E, R, L_{ER}\}$, for $e_x, e_y \in E$, there are $s_i \in S, r_j \in R, l_{xi} \in L_{ES}, l_{yi} \in L_{ES}, l_{sj} \in L_{ER}$, and $l_{rj} \in L_{ER}$, and then competition paths can be defined as $Compath(e_x \rightarrow s_i \rightarrow e_y)$ and $Compath(e_x \rightarrow r_j \rightarrow e_y)$. Obviously, if the above paths exist, there are competition paths $Compath(e_y \rightarrow s_i \rightarrow e_x)$ and $Compath(e_y \rightarrow r_j \rightarrow e_x)$.

This study uses above definition and constructed bipartite graphs to identify competition paths between paired enterprises, as shown in Table 7.

Table 7: Competition paths identification between enterprises (an illustrative example).

Competition paths of skills demand	Example	Competition paths of resources supply	Example
$e_1 \rightarrow s_i \rightarrow e_2$	$e_1 \rightarrow Product\ marketing \rightarrow e_2$ $e_1 \rightarrow Website\ OM \rightarrow e_2$	$e_1 \rightarrow r_j \rightarrow e_2$	$e_1 \rightarrow Information\ system \rightarrow e_2$ $e_1 \rightarrow Marketing\ tactics \rightarrow e_2$
$e_1 \rightarrow s_i \rightarrow e_3$	$e_1 \rightarrow Data\ analysis \rightarrow e_3$	$e_1 \rightarrow r_j \rightarrow e_3$	$e_1 \rightarrow Clients\ database \rightarrow e_3$
$e_2 \rightarrow s_i \rightarrow e_3$	$e_2 \rightarrow Video\ editing \rightarrow e_3$	$e_2 \rightarrow r_j \rightarrow e_3$	$e_2 \rightarrow Patent\ technology \rightarrow e_3$

Asymmetric competitiveness degrees calculation based on the above competition paths

Competitiveness degree is a relative measure of dynamic competition in a particular market environment, defined by a given enterprises set. Once an enterprises set is defined, competitiveness degree of every paired enterprise can be calculated (Wei et al., 2016). Therefore, competitiveness degree is a kind of reflect and measure to competitive degree, and the two concepts are not distinguished in this study. To calculate asymmetric competitiveness degrees, the first step is to determine intensity of skills demand or resources supply for the enterprises included in the competition paths, which can be defined as follows:

Definition 3: Given $l_{xi} \in L_{ES}$ and $l_{yj} \in L_{ER}$, the e_x ' demand for s_i can be defined as $|e_x s_i|$, and supply for r_j can be defined as $|e_x r_j|$. For s_i and r_j , the probability of co-occurrence with e_x can be defined as $p(e_x \rightarrow s_i) = |e_x s_i| / |e_x|$ and $p(e_x \rightarrow r_j) = |e_x r_j| / |e_x|$.

For e_x , the probability of co-occurrence with s_i and r_j can be defined as $p(s_i \rightarrow e_x) = |s_i e_x| / |s_i|$ and $p(r_j \rightarrow e_x) = |r_j e_x| / |r_j|$.

Where $|e_x s_i|$ and $|e_x r_j|$ reflect intensity of skills and resources. $|e_x|$ is intensity sum of skills demanded for or resources supplied to e_x compete with other enterprises. In skills demand, $|e_x|$ reflects the extent to which e_x considers s_i as an important skill for itself. This is identification of specific skill from skills pool owned by the focus enterprise, in turn this reflects enterprise' demand intensity for that skill. Similarly, in resources supply, $|e_x|$ can reflect the extent to which e_x considers r_j as an important resource for its own. This is identification of specific resource from resources pool owned by the focus enterprise, in turn this reflects enterprise' supply intensity for that resource. $|s_i|$ is demand intensity sum of all enterprises for this skill, and reflects the extent to which e_x considers s_i may be needed by itself, that is, the probability of s_i being demanded as one of many skills. This is identification of specific enterprise from enterprises pool that demands the skill, and reflects demand

intensity for that skill. Similarly, $|r_j|$ is supply intensity sum of all enterprises for this resource, and reflects the extent to which e_x considers r_j to be own resource capability, that is, the probability of S_i being possessed as one of many resources. This is identification of specific enterprise from enterprises pool that supplies the resource, and reflects supply intensity for that resource. See Table 8 for details.

Table 8: Skills demand or resources supply intensity (an illustrative example).

Skills Demand Intensity $ e_x S_i $	Overall Skills Demand Intensity $ e_x $	Resources Supply Intensity $ e_x r_j $	Overall Resources Supply Intensity $ e_x $	
$ e_1 \text{Product marketing} =12$ $ e_1 \text{Data analysis} =10$ $ e_1 \text{Website OM} =9$	$ e_1 =12+10+9=31$	$ e_1 \text{Information system} =9$ $ e_1 \text{Clients database} =8$ $ e_1 \text{Marketing tactics} =8$	$ e_1 =9+8+8=25$	
$ e_2 \text{Product marketing} =15$ $ e_2 \text{Website OM} =11$ $ e_2 \text{Video editing} =10$		$ e_2 \text{Information system} =9$ $ e_2 \text{Marketing tactics} =12$ $ e_2 \text{Patent technology} =11$		$ e_2 =9+12+11=32$
$ e_3 \text{Data analysis} =13$ $ e_3 \text{Video editing} =12$		$ e_3 \text{Clients database} =11$ $ e_3 \text{Patent technology} =13$		$ e_3 =11+13=24$
	$ \text{Product marketing} = 12 + 15 = 27$		$ \text{Information system} = 9 + 9 = 18$	
$ S_i $	$ \text{Data analysis} = 10 + 13 = 23$	$ r_j $	$ \text{Clients database} = 8 + 11 = 19$	
	$ \text{Website OM} = 9 + 11 = 20$		$ \text{Marketing tactics} = 8 + 12 = 20$	
	$ \text{Video editing} = 10 + 12 = 22$		$ \text{Patent technology} = 11 + 13 = 24$	

Next, to calculate asymmetric competitiveness degrees for a particular skill demand between paired enterprises, it can be defined as follows:

Definition 4: For e_x and e_y , according to $Compath(e_x \rightarrow s_i \rightarrow e_y)$ and $Compath(e_x \rightarrow r_j \rightarrow e_y)$, the competitiveness degree between enterprises in S_i and r_j can be defined as $Comps_i(e_x, e_y) = p(e_x \rightarrow s_i) \times p(s_i \rightarrow e_y)$, $Comps_i(e_y, e_x) = p(e_y \rightarrow s_i) \times p(s_i \rightarrow e_x)$, and $Compr_j(e_x, e_y) = p(e_x \rightarrow r_j) \times p(r_j \rightarrow e_y)$, $Compr_j(e_y, e_x) = p(e_y \rightarrow r_j) \times p(r_j \rightarrow e_x)$.

Definition 3 and 4 involve overall skills or resources of different enterprises. This reflects range and extent to which different enterprises' skills demand or resources supply. For example, if enterprise focuses on a small number of skills or resources, it will be more competitive in a particular skill or resource. On the other hand, enterprise focuses on a lot of skills or resources, then it will be less competitive in a certain skill or resource. Then, the asymmetric competitiveness degrees between enterprises can be observed, and are very common in real world. When $|e_x|$ and $|e_y|$ is equal, the competitiveness degrees are theoretically equal, which is virtually non-existent. Finally, the overall asymmetric competitiveness degrees of paired enterprises can be calculated as follows:

Definition 5: For e_x and e_y , the overall asymmetric competitiveness degree can be defined as $Comp(e_x, e_y) = \frac{1}{2} \sum_{s_i \in S} Comps_i(e_x, e_y) + \frac{1}{2} \sum_{r_j \in R} Compr_j(e_x, e_y)$, which has a range of [0,1], and $e_y \in E$. Similarity, the asymmetric competitiveness degree is $Comp(e_y, e_x) = \frac{1}{2} \sum_{s_i \in S} Comps_i(e_y, e_x) + \frac{1}{2} \sum_{r_j \in R} Compr_j(e_y, e_x)$.

This definition reflects overall asymmetric competitiveness degrees of enterprises in multiple competition paths. For e_x , in skills demand and resources supply, the overall asymmetric competitiveness degrees of focus enterprise with all other enterprises in enterprises set is 1, which can be proved mathematically. This also describes an actual context: for a given enterprises set, the scope of competition analysis is defined as each enterprise in the set.

Based on the above definition and demand or supply intensity, this study calculates skills and resources asymmetric competitiveness degrees of paired enterprises, and then gets overall asymmetric competitiveness degrees, as shown in Table 9.

Table 9: The asymmetric competitiveness degrees between enterprises (an illustrative example).

Enterprises	Competitiveness Degree of Skills Demand between Enterprises			Competitiveness Degree of Resources Supply between Enterprises			Overall Competitiveness Degree between Enterprises		
	e_1	e_2	e_3	e_1	e_2	e_3	e_1	e_2	e_3

e_1	0.4429	0.3747	0.1823	0.4427	0.3720	0.1853	e_1	0.4428	0.3734	0.1838
e_2	0.3227	0.5258	0.1515	0.2906	0.5232	0.1862	e_2	0.3067	0.5245	0.1689
e_3	0.2261	0.2182	0.5557	0.1930	0.2483	0.5588	e_3	0.2095	0.2332	0.5572

Experiment of the above Competition Analysis Method

This part uses real enterprise recruitment texts and above methods to analyze competition. And, this is to verify validity and advantages of the above method, as well as provide support for enterprise competition strategy and evaluation based on the results.

Competition analysis experiment

Details of the 10 representative enterprises are shown in Table 10. They are the 10 enterprises with the highest number of recruitment texts in IT and finance industry respectively during data collection period. At present, these two industries not only have a large increase in talents demand (Hu *et al.*, 2019), but also have a cross-industry combination in technologies and functions (Du *et al.*, 2019). The competition analysis between them can check the method validity, examine inter- and exter-industry competition, as well as calculate asymmetric competitiveness degrees between enterprises. The algorithm of competition analysis is shown in Table 11, and the asymmetric competitiveness degrees matrix is shown in Table 12.

Table 10: Descriptive information of ten enterprises.

No	Industry	Recruitment Information Quantity	Size	Nature	Samples of Recruitment Position
1	IT/EC	293	500-1000	JVE	Foreign exchange trader, financial data analysts
2	IT/EC	288	500-1000	JVE	Customer servicer, office clerk
3	Finance/Investment	283	500-1000	PE	Financial data analysts, foreign exchange trader
4	IT/EC	255	500-1000	JVE	EC operation specialist, data analysis specialist
5	Finance/Investment	226	100-500	JVE	Financial transactions specialist, data analysis specialist
6	Finance/Investment	191	100-500	JVE	Financial transactions specialist, financial analysts
7	IT/EC	183	500-1000	FIE	Data analyst, foreign exchange trader
8	IT/EC	181	500-1000	FIE	Data analysis specialist, PA specialist
9	Finance/Investment	180	100-500	PE	Financial analysts, data analyst
10	Finance/Investment	177	50-150	PE	Foreign exchange trader, financial transactions specialist

Tips: EC means electronic commerce; JVE means joint venture enterprise; PE means private enterprise; FIE means foreign-invested enterprise; PA means performance appraisal.

Table 11: The algorithm of competition analysis.

Algorithm: Competition analysis process between companies

```

Input: Company entities set  $E$ , job descriptions  $E\_JD$ , company profiles  $E\_EP$ , skills demand indicators  $S$ , resource base indicators  $R$ 
for each  $e$  in  $E$  do
  for each  $e\_jd$  in  $E\_JD$  do
    # Extracting skills demand indicators  $e\_s_i$ 
    if  $e\_s_i \in S$  then
       $e\_s_i = e\_s_i + 1$ 
    end if
  end for
  # " for each  $e\_ep$  in  $E\_EP$  do " and "Extracting resource base indicators  $e\_r_j$  ", same as above
end for
 $E\_S = \{e \in E | e\_s\}$ ,  $E\_R = \{e \in E | e\_r\}$  # Constructing companies-skills and companies-resources bipartite graphs
for each  $e_x$  in  $E$  do
  for each  $e_y$  in  $E$  do
    if  $\exists s_i \in (e_x\_s \cap e_y\_s)$  then
       $Comps(e_x, e_y) = skills\_demand(e_x, e_y, s_i)$  # Calculating the competitiveness degree of skills demand about  $e_x$  to  $e_y$ 
    end if
  end for
end for

```

" if $\exists r_j \in (e_x - r \cap e_y - r)$ then" and "Calculating the competitiveness degree of resource base about e_x to e_y ", same as above
 end for
 end for

$$Comp(e_x, e_y) = \frac{1}{2} Comps(e_x, e_y) + \frac{1}{2} Compr(e_x, e_y)$$
 # Calculating the competitiveness degree about e_x to e_y
 # " for each e_y in E do" and "Calculating the competitiveness degree about e_y to e_x ", same as above
 Output: $Comp(e_x, e_y)$, $Comp(e_y, e_x)$ # Obtaining the asymmetric competitiveness degrees matrix between paired companies

Table 12: The asymmetric competitiveness degrees matrix between enterprises.

E	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
e_1	0.2266	0.0330	0.2189	0.0218	0.1060	0.1178	0.0552	0.0557	0.1159	0.0491
e_2	0.0132	0.8650	0.0298	0.0099	0.0101	0.0290	0.0084	0.0149	0.0125	0.0072
e_3	0.1492	0.0617	0.2829	0.0314	0.0757	0.1037	0.0529	0.0735	0.1293	0.0398
e_4	0.0663	0.0297	0.1617	0.3167	0.0738	0.0297	0.0530	0.1426	0.0899	0.0367
e_5	0.1082	0.0165	0.1177	0.0248	0.3869	0.0926	0.0641	0.0209	0.0755	0.0929
e_6	0.0870	0.0536	0.1032	0.0102	0.0588	0.4974	0.0690	0.0256	0.0538	0.0414
e_7	0.1036	0.0154	0.1420	0.0338	0.1243	0.1888	0.2332	0.0502	0.0679	0.0408
e_8	0.0559	0.0170	0.1109	0.0494	0.0159	0.0296	0.0232	0.6096	0.0538	0.0346
e_9	0.1547	0.0378	0.2580	0.0359	0.0971	0.1013	0.0517	0.0719	0.1451	0.0466
e_{10}	0.1075	0.0100	0.0942	0.0170	0.2576	0.1242	0.0435	0.0421	0.0750	0.2288

Content validity of the competition analysis evaluation index systems

In this part, three experts in EC field in two universities were invited to evaluate the index systems, that is, to test content validity of the index systems. Questionnaire design: (1) 50 skill indicators and 50 resource indicators were randomly selected from the index systems. (2) The question "Do you think this indicator(term) is a valid indicator of skills demand (or resources supply) of enterprise?" and the options "Validity, recommended to be retained" and "Not recommended to be retained" were set. Survey process: (1) The paper questionnaire including above 100 skill and resource indicators were provided to three experts, who were given definitions of skills demand and resources supply and understood the meaning of question. If the experts still have questions, researchers continue to answer in-depth. The experts could fill in questionnaire until were no longer in doubt. (2) After three working days, the questionnaire was retrieved. Data processing: Content validity ratio CVR was calculated for each indicator, i.e., $CVR = (Ne - N/2) / (N/2)$. Where, Ne is the number of experts who chose "Validity, recommended to be retained", and N is the total number of experts, as well as value range of CVR is [-1,1]. If CVR is positive, the indicator should be retained, and if negative, should not be retained (Wei & Huang, 2015; Yu & Wen, 2011). The CVR of 100 indicators were all positive, which indicated the indicators selected randomly were all validity.

Results and Visualization

Taking e_1 in the asymmetric competitiveness degrees matrix as an example, the following competition decision analysis can be carried out to provide decision-making recommendations for competition practice. First, according to the asymmetric competitiveness degrees matrix, the asymmetric competition relations between focal company and others can be compared, and then the set of competitors with relative (dis)advantages can be delimited. For example, the focal company has equal competitiveness degrees with e_1 and e_8 , and has advantages when facing e_2 , e_3 , and e_6 , while it has disadvantages when facing e_4 , e_7 , e_9 , and e_{10} . Then, according to the set of competitors, focal company can adopt various competition strategies to consolidate or improve its market share or position in industry. In addition, according to the various competition strategies, and combining with concrete competition paths, focal company can adjust or determine future competition focus. For example, when facing e_3 , the focal company can take active strategies in such fields as financial investment, marketing, data analysis and security control, stock trading, and international financial trade, and so on. When facing e_9 , it can take strategic actions in such areas as online marketing, customer service, and online platform building, and so on, which will consume more resources and last longer. Finally, considering the identified skills or resources focus, focal company can evaluate their monopoly fields in skills or resources and competing solutions. For example, it can integrate and build internal and external capabilities to respond to a rapidly changing environment by proactively perceiving, seizing opportunities, and making adaptive changes (Teece et al., 1997).

In order to enhance the interpretability about the overall competitive situation, the asymmetric competitiveness degrees matrix is taken as input of Social Network Analysis (SNA), and the centralities of each company are calculated. The results are illustrated in Figure 5. Concretely, horizontal and vertical axis respectively represent outdegree and indegree, that is, the competitiveness degree of the focal company to other companies, and other companies to the focal company. In the context of

this study, we, based on different competing entity perspectives, named the outdegree as competitive attack, while named the indegree as competitive threat.

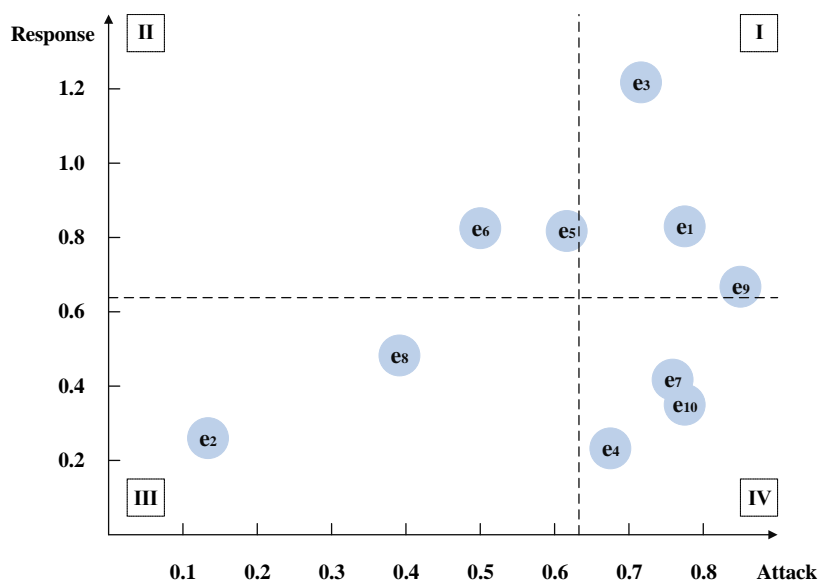


Figure 5: The overall competitive situation based on asymmetric competitiveness degrees.

According to average outdegree and average indegree of companies set, the overall competitive situation can be divided into four quadrants. In quadrant I, both outdegree and indegree are above the average. When companies are faced with fierce competition (e.g., e_1 , and e_3), they are distributed in quadrant I, which means both the threat from the focal company to other companies and the threat from other companies to the focal company are significant. In quadrant II, outdegree is below the average and indegree is above the average. When firms lack a competitive advantage (e.g., e_6), they are distributed in quadrant II, which means that other firms are more of a threat to the focal firm than the focal firm is to other firms. In quadrant III, outdegree and indegree are below the average. When companies are exclusive (i.e., e_2 and e_8), they are usually distributed in quadrant III, which means that the two threats are small. In quadrant IV, outdegree is above the average and indegree is below the average. When firms have a competitive advantage (e.g., e_4), they are distributed in quadrant IV, which means that the focal firm is more of a threat to other firms than other firms are to the focal firm.

Baseline Model Comparison

Given that there is no algorithm proposed for identifying firm competitive relationships from recruitment text data, we compare the results of firm competitive relationship obtained from the analysis of recruitment texts using classical coupling index method and different text classification and matching methods with the approach proposed in this article to validate the advantages of our proposed approach.

The coupling index has been proposed to reflect correlations among subjects. In a given entities set, coupling index is the ratio of the number of shared technologies over the sum of technologies of two paired entities (Lo, 2008). And, few extant studies suggested the coupling index can reflect similarity between subjects (Hong & Li, 2013). Therefore, this study uses coupling index to calculate similarity of skills demand and resource base of the ten companies in the experiment.

By comparing the results of coupling index, we found: (1) the coupling index method only measures the similarity of skills and resources between companies but cannot conduct asymmetric competition analysis; (2) the coupling index method only considers skills and resources set when calculating competitiveness degree, and has some deficiencies in visualization of competition relations, identification of competition paths, and analysis of competition focus. However, competitiveness degree between paired companies is rarely equal, and asymmetric competitiveness degrees are more common.

Second, we also use TF-IDF and BERT methods as baseline, which can be treated as a simplified version of our method. Concretely, the job descriptions and company profiles are represented by using these two methods, and the competitiveness degrees matrix among companies are obtained.

Also, we chose two state-of-the-art text classification and matching methods as the baseline models. One is Long Short-Term Memory (LSTM) network, the other is Convolutional Neural Network (CNN).

The results of the focal company overall competitive situation based on SNA are illustrated in Figure 6. The Figure shows that the modified bipartite graph approach (Modified BGA) can divide the competitors into different sections, which the focal company is at an advantage or a disadvantage or the degrees of mutual competitiveness are equal, by revealing the asymmetric competitiveness degree. For example, facing with e_2 , e_3 , and e_6 , e_1 has a competitive advantage, facing with e_4 , e_7 , e_9 ,

and e_{10} , e_1 has a competitive disadvantage, and facing with e_5 and e_8 , the degree of threat between e_1 and them is equal. However, the other five methods (*i.e.*, Coupling Index, TF-IDF, BERT, LSTM, and CNN) can only get the competitiveness degree of the focal company compared with the average competitiveness degree between focal and all other companies, by revealing the symmetrical competitiveness degree. Take the coupling index method as an example, which is similar to the TF-IDF, BERT, LSTM, and CNN, the competitiveness degrees between e_1 and e_2 , e_4 , e_8 , e_{10} are lower than the average, but the competitiveness degrees between e_1 and e_3 , e_5 , e_6 , e_7 , e_9 are higher than the average. On top of this, the five baseline models are unable to disclose the overall competitive situation.

We use the error-rank ratio (ERR) to quantify the method accuracy. The assessment of competitiveness rank relies on using three experts to evaluate and rate ten companies based on their relevance and representativeness to IT and finance domain. A questionnaire was designed with (1) information of 10 companies; (2) two questions "Please rank the nine companies (*i.e.*, e_2 to e_{10}) based on threat degree of focal company (*i.e.*, e_1) to them" and "Please rank the nine companies (*i.e.*, e_2 to e_{10}) based on threat degree of them to focal company (*i.e.*, e_1)". The ranks of three experts were respectively averaged to obtain two final competitiveness ranks, which represent competitive attack and competitive response of focal company. The ERR of variety

methods was calculated, *i.e.*,
$$ERR = \sqrt{\sum_{j=2}^{10} (Rank_{KE,j} - Rank_{ME,j})^2}$$
, where $Rank_{KE,j}$ is the rank of expert evaluation, and $Rank_{ME,j}$ is the rank of methods (*i.e.*, MBGA, CI, TF-IDF, BERT, LSTM, and CNN). The competitiveness rank and ERR are shown in Table 13. The result indicates the MBGA method not only provides insight into the asymmetric competitiveness degrees, but has the lowest ERR.

CONCLUSION AND DISCUSSION

Online recruitment services offer a viable solution to the difficult situation between employers and job seekers during the 2019 Coronavirus Disease (COVID-19) global pandemic by providing timely and effective job search and hiring information. The rapid growth of online recruitment services has also accumulated a large amount of recruitment text data which contain rich information regarding employers and job seekers. Although the existing studies have proposed a myriad of strategies to study competitor analysis and inter-firm rivalry from different data sources (*e.g.*, site-centric data, in-links and out-links of websites, etc.), limited efforts have been devoted to implementing competitor analysis from recruitment text data. In this article, we proposed a feasible method framework to analyze companies' competitors and asymmetric competitiveness degrees from the massive unstructured recruitment texts. The key idea is to exploit the wealth of information in the job description and company profile in the recruitment text. We first proposed semantic representations for both job descriptions and company profiles and constructed skills demand and resource base competition evaluation indicators from the job descriptions and company profiles. Then, we used the competition evaluation index system and a modified bipartite graph method to display competition relations. Finally, we used the results of bipartite graphs to identify competition paths between paired companies and calculated asymmetric competitiveness degrees. Based on a real-world dataset, we clearly showed the effectiveness and interpretability of the proposed method compared to some baseline models.

This research has two theoretical significances. First, this study contributes to the massive recruitment texts mining literature by enriching method framework of competitive intelligence. The method framework uses Word2Vec to extract competition information of skills demand and resources supply, and uses K-means to construct the evaluation index systems, then gets bipartite graphs of enterprises-skills and enterprises-resources to quantify asymmetric competitiveness degrees. Second, this study contributes to the market and resource competition theories literature by enriching research context of data source used in competition analysis. Based on theories of market commonality and resource similarity, this study finds recruitment texts has competition analysis attributes.



Figure 6: The focal company overall competitive situation based on SNA using variety methods.

Table 13: The competitiveness rank and error-rank ratio (ERR) based on e_1 using variety methods.

Co.	Competitive Attack							Competitive Response						
	EE	MBGA	CI	TF-IDF	BERT	LSTM	CNN	EE	MBGA	CI	TF-IDF	BERT	LSTM	CNN
e_2	2	2	8	8	8	2	4	8	8	8	8	8	2	4
e_3	8	5	2	2	2	8	6	2	2	2	2	2	8	6
e_4	5	8	6	6	4	4	2	6	4	6	6	4	4	2
e_5	4	4	4	5	6	5	5	9	9	4	5	6	5	5
e_6	9	7	5	4	9	1	9	4	6	5	4	9	1	9
e_7	6	6	7	9	5	7	7	5	5	7	9	5	7	7
e_8	7	9	9	7	7	9	8	3	3	9	7	7	9	8
e_9	1	1	1	3	1	6	3	1	7	1	3	1	6	3
e_{10}	3	3	3	1	3	3	1	7	1	3	1	3	3	1
ERRa/r	-	5.10	9.70	10.77	8.83	9.80	5.29	-	8.94	9.06	9.38	8.37	13.49	12.57
Overall	-	14.04	18.75	20.15	17.20	23.29	17.86	-	-	-	-	-	-	-

Notes: Co. (companies); EE (Expert Evaluation); MBGA (Modified Bipartite Graph Approach); CI (Coupling Index); ERRa/r (error-rank ratio of attack or response); Overall (overall error-rank ratio, which is the sum of "Competitive Attack" and "Competitive Response").

This study has three practical implications. First, this study is beneficial to enterprises to rapidly identify competition focus and quantify competitiveness degree through recruitment texts. Quickly extracting competitive knowledge and transforming into decision-making basis can help enterprises to analysis competitors, use appropriate strategies, and play a more influential,

rather than passive or reactive, role in competition. The competitiveness degrees help managers understand how competitors tend to switch skills and resources dynamically, and obtain further insights in analyzing the dynamics of skills demand and resources supply of competitors. Second, this research is beneficial to recruitment platform to grasp enterprises overall development, and then to promote effective matching of talents. By understanding enterprises reality and forming a good interactive environment, the recruitment platform can attract enterprises to register and post positions, and then attract applicants to register and upload resumes, finally improve success rate of talents and enterprises matching. Third, this research is helpful for applicants to grasp enterprises development focus and trends, so as to quickly select out enterprises, which match their own professional skills and career.

This study has some limitations that future research could address. First, we used word2Vec as input feature instead of the Bag-of-Words as input features. A potential limitation of using this approach is that the pre-trained word vectors may not be sufficient to describe the semantic features of the recruitment text data, which can affect the accuracy of the index systems construction for competitive analysis. Future research could attempt to use state-of-the-art deep learning text processing methods for semantic-based analysis, such as the deep graph CNN model and hierarchical attention network to improve the accuracy of the index systems. Besides, the system of competition analysis indicators established in this study covers as many levels and dimensions as possible, but may still exclude the existence of some other valid indicators. Future studies can expand the adaptability based on the indicator system of this study, taking into account the specific situation. Second, in addition to the accuracy of the text semantic algorithm, the quality of the evaluation indicators in this study also depends heavily on the manual inspection of features or key phrases by domain experts. However, manual inspection leads to high costs and inefficient, inaccurate and subjective judgments. Thus, future research is encouraged to develop new approaches to screen indicators without relying on human judgment. Finally, this study mainly focuses on the Chinese recruitment platform texts, and does not involve the non-Chinese recruitment texts. Future research could consider investigating in different linguistic settings and national cultural contexts to discover and deepen the applicability of the study.

ACKNOWLEDGMENT

This work is supported by the Social Science Foundation of Sichuan Province (SC22B144).

REFERENCES

- An, L., Hu, J., & Li, G. (2019). A method of identifying comment opinion leaders based on topic consistency and emotional support. *Journal of Management Science*, 32(01), 3-13. <https://doi.org/10.3969/j.issn.1672-0334.2019.01.001> (in Chinese).
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120. <https://doi.org/10.1177/014920639101700108>
- Chakraborti, S., & Dey, S. (2014). Multi-document text summarization for competitor intelligence: A methodology. 2014 2nd International Symposium on Computational and Business Intelligence, 97-100. <https://doi.org/10.1109/ISCBI.2014.28>
- Chen, M. J. (1996). Competitor analysis and interfirm rivalry: Toward a theoretical integration. *Academy of Management Review*, 21(1), 100-134. <https://doi.org/10.5465/amr.1996.9602161567>
- Choi, J., Tosyali, A., Kim, B., Lee, H. S., & Jeong, M. K. (2022). A novel method for identifying competitors using a financial transaction network. *IEEE Transactions on Engineering Management*, 69(04), 845-860. <https://doi.org/10.1109/TEM.2019.2931660>
- Dai Zefan, Zhu Yifan, & Xu Xin. (2018). A competitor identification method based on user coincidence. *Information Studies: Theory & Application*, 41(09), 57-62. <https://doi.org/10.16353/j.cnki.1000-7490.2018.09.010> (in Chinese)
- Deng Xinming, Long Liangzhi, Long Xianyi, & Ye Zhen. (2017). Research on the correlation between competitor analysis and market, nonmarket attacks. *Management Review*, 29(11), 129-147. <https://doi.org/10.14120/j.cnki.cn11-5057/f.2017.11.012> (in Chinese)
- Du Jun, Han Zihui, & Jiao Yuanyuan. (2019). Evolution path and realization path of profit model for internet financial services: A case study of JD supply chain finance. *Management Review*, 31(8), 277-294. <https://doi.org/10.14120/j.cnki.cn11-5057/f.2019.08.024> (in Chinese)
- Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management*, 71(01), 19-32. <https://doi.org/10.1016/j.ijhm.2017.09.004>
- Gardiner, A., Aasheim, C., Rutner, P., & Williams, S. (2018). Skill requirements in big data: A content analysis of job advertisements. *Journal of Computer Information Systems*, 58(4), 374-384. <https://doi.org/10.1080/08874417.2017.1289354>
- Gupta, A., Kumar, A., Grewal, R., & Lilien, G. L. (2019). Within-seller and buyer-seller network structures and key account profitability. *Journal of Marketing*, 83(1), 108-132. <https://doi.org/10.2139/ssrn.2933839>
- Gutt, D., Herrmann, P., & Rahman, M. S. (2019). Crowd-driven competitive intelligence: Understanding the relationship between local market competition and online rating distributions. *Information Systems Research*, 30(3), 980-994. <https://doi.org/10.1287/isre.2019.0845>
- Hong Yong, & Li Yingmin. (2013). Visualization of technological similarity analysis for enterprises based on patent coupling approach. *Studies in Science of Science*, 31(07), 1013-1021. <https://doi.org/10.16192/j.cnki.1003-2053.2013.07.008> (in Chinese)

- Hu Beibei, Wang Shengguang, & Duan Yuchang. (2019). An analysis of the new Techno-Economic paradigm caused by the Internet. *Studies in Science of Science*, 37(04), 582-589. <https://doi.org/10.16192/j.cnki.1003-2053.2019.04.002> (in Chinese)
- Jantunen, A., Puumalainen, K., Saarenketo, S., & Kyläheiko, K. (2005). Entrepreneurial orientation, dynamic capabilities and international performance. *Journal of International Entrepreneurship*, 3(3), 223-243. <https://doi.org/10.1007/s10843-005-1133-2>
- Liu, Y., Qian, Y., Jiang, Y., & Shang, J. (2020). Using favorite data to analyze asymmetric competition: Machine learning models. *European Journal of Operational Research*, 287(2), 600-615. <https://doi.org/10.1016/j.ejor.2020.03.074>
- Lo, S. C. (2008). Patent coupling analysis of primary organizations in genetic engineering research. *Scientometrics*, 74(1), 143-151. <https://doi.org/10.1007/s11192-008-0110-7>
- Long Haiquan, Lv Benfu, Peng Geng, & Zhao Tianbo. (2010). Research of Internet companies' core resources and capabilities: Based on the perspective of value creation. *Chinese Journal of Management Science*, 18(1), 161-167. <https://doi.org/10.16381/j.cnki.issn1003-207x.2010.01.021> (in Chinese)
- Ma Yongfang. (2015). Competitor identification based on market commonality and resource similarity (Master dissertation, Tsinghua University, Beijing, China). (in Chinese)
- Ma, Z., Pant, G., & Sheng, O. R. (2011). Mining competitor relationships from online news: A network-based approach. *Electronic Commerce Research and Applications*, 10(4), 418-427. <https://doi.org/10.1016/j.elerap.2010.11.006>
- Mohammed, I., Guillet, B. D., & Law, R. (2014). Competitor set identification in the hotel industry: A case study of a full-service hotel in Hong Kong. *International Journal of Hospitality Management*, 39(05), 29-40. <https://doi.org/10.1016/j.ijhm.2014.02.002>
- Nam, H., Joshi, Y. V., & Kannan, P. K. (2017). Harvesting brand information from social tags. *Journal of Marketing*, 81(4), 88-108. <https://doi.org/10.1509/jm.16.0044>
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543. <https://doi.org/10.1287/mksc.1120.0713>
- Pant, G., & Sheng, O. R. (2015). Web footprints of firms: Using online isomorphism for competitor identification. *Information Systems Research*, 26(1), 188-209. <https://doi.org/10.1287/isre.2014.0563>
- Peng, Y. S., & Liang, I. C. (2016). A dynamic framework for competitor identification: A neglecting role of dominant design. *Journal of Business Research*, 69(5), 1898-1903. <https://doi.org/10.1016/j.jbusres.2015.10.076>
- Porac, J. F., & Thomas, H. (1990). Taxonomic mental models in competitor definition. *Academy of Management Review*, 15(2), 224-240. <https://doi.org/10.5465/amr.1990.4308155>
- Qiao, D., Zhang, J., Wei, Q., & Chen, G. (2017). Finding competitive keywords from query logs to enhance search engine advertising. *Information & Management*, 54(4), 531-543. <https://doi.org/10.1016/j.im.2016.11.003>
- Ringel, D. M., & Skiera, B. (2016). Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science*, 35(3), 511-534. <https://doi.org/10.1287/mksc.2015.0950>
- Smith, D., & Ali, A. (2014). Analyzing computer programming job trend using web data mining. *Issues in Informing Science and Information Technology*, 11(1), 203-214. <https://doi.org/10.28945/1989>
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509-533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO2-Z)
- Turrell, A., Speigner, B. J., Djumalieva, J., Copple, D., & Thurgood, J. (2019). Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings. *National Bureau of Economic Research*, 19(05), 1-36. <https://doi.org/10.3386/w25837>
- Wang Yi, Chen Jin, & Xu Qingrui. (2000). Core competence of corporation: Theory sources and logical structure. *Journal of Management Sciences in China*, 4(03), 24-32+43. <https://doi.org/10.3321/j.issn:1007-9807.2000.03.004> (in Chinese)
- Wei Jiang, & Huang Xue. (2015). The innovation capability evaluation index system for high-tech services. *Science Research Management*, 36(12), 9-18. <https://doi.org/10.19571/j.cnki.1000-2995.2015.12.002> (in Chinese)
- Wei, Q., Qiao, D., Zhang, J., Chen, G., & Guo, X. (2016). A novel bipartite graph based competitiveness degree analysis from query logs. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11(2), 1-25. <https://doi.org/10.1145/2996196>
- Wu, X., Xu, T., Zhu, H., Zhang, L., Chen, E., & Xiong, H. (2019). Trend-aware tensor factorization for job skill demand analysis. In S. Kraus (Ed.), *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019* (pp. 3891-3897). (IJCAI International Joint Conference on Artificial Intelligence; Vol. 2019-August). International Joint Conferences on Artificial Intelligence. <https://doi.org/10.24963/ijcai.2019/540>
- Xu, X., Dutta, K., Datta, A., & Ge, C. (2018). Identifying functional aspects from user reviews for functionality-based mobile app recommendation. *Journal of the Association for Information Science and Technology*, 69(2), 242-255. <https://doi.org/10.1002/asi.23932>
- Yan Weiwei, & Ma Xiaoxiao. (2019). The organizing behavior rules and references for web audio resources collection. *Journal of Modern Information*, 39(05), 74-79+88. <https://doi.org/10.3969/j.issn.1008-0821.2019.05.010> (in Chinese)
- Yu Rongjian, & Wen Kai. (2011). Uncovering the "Black Box" of GVC governance: Structure, model, mechanism and its influence-A cross-case study based on 12 OEM relationship between Zhejiang Merchants. *Management World*, (08), 142-154. (in Chinese)
- Zhang Yin, Zhao Wenhui, Bao Hengyue, Li Yajian, & Zhou Keqiang. (2017). Data mining of online recruitment information based on K-means and correlation analysis. *Software Engineering*, 20(05), 10-14. (in Chinese)

Zhong, X., Chen, G., Sun, L., & Liu, L. (2018). Data-driven competitor identification using text mining and word network analysis. In *Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018)* (pp. 461-468).