

# Analyzing the Personality of Party and Government Leaders: A LDA Topic Model

Dengke Yu<sup>1</sup> & Lina Le<sup>1</sup>

<sup>1</sup> School of Public Policy and Administration, Nanchang University, Nanchang Jiangxi 330031, China

Correspondence: Dengke Yu, School of Public Policy and Administration, Nanchang University, Nanchang Jiangxi 330031, China. E-mail: 304299943@qq.com

Received: November 04, 2023 Accepted: November 15, 2023 Online Published: December 5, 2023

doi:10.5430/bmr.v13n1p50

URL: <https://doi.org/10.5430/bmr.v13n1p50>

*The research was financed by National Natural Science Foundation of China, grant number 71962021 and 72362028.*

## Abstract

**Purpose/Significance:** Personalities of party and government leaders play a pivotal role in shaping their preferences, attitudes, and conduct in the context of strategic decision-making. The analysis of personality carries substantial implications for the assessment and selection of leadership cadres. The study endeavors to introduce an innovative method for scrutinizing the personalities of party and government leaders, leveraging the technological topics generated through a Latent Dirichlet Allocation (LDA) model.

**Methods/Procedures:** Anchored in the theory of personality behavior, the study posits that discernible topics can be extracted from news reports that document the political activities of party and government leaders. These discerned topics can be employed to make inferences about their personalities. Consequently, we systematically amassed news reports chronicling the political engagements of 62 party and government leaders spanning 31 provinces, municipalities, and autonomous regions in mainland China, covering their tenures in office up to 2021. Based on the LDA model, we conducted an analysis to uncover the dimensional structure of their personalities, incorporating techniques such as topic clustering and common factor extraction.

**Results/Conclusions:** The study effectively identified six distinct personalities that correspond with Hollander's views on occupational personality. The development gave rise to a theory that specifically addresses the personalities exhibited by party and government leaders. Spatial analysis validated the presence of a spatial aggregation effect, underscoring the validity of our framework. Our study can provide implications for the training, evaluation, and designation of party and government leaders in the future.

**Keywords:** party and government leader, personality, LDA model, spatial analysis

## 1. Introduction

Party and government leaders serve as the primary decision-making and implementation body in the process of advancing the development of national governance system and enhancing regions' governance capacities. They have the pivotal force determining the feasibility of implementing a country's top-level institutional design (Liu et al., 2022). In tandem with the accelerated economic transformation and the modernization of social governance in China, there is a discernible departure from sole reliance on GDP performance evaluation in the selection and appointment of party and government leaders. Instead, greater emphasis is placed on the principles of "fitness for duty" and "fitness for virtue" (Shang, 2018). In academic field, research on the management of party and government leaders and cadres predominantly focuses on their political backgrounds, educational qualifications, talents, and performance (Ji & Zhang, 2022; He, 2020). However, the personalities of party and government leaders, which play a crucial role in shaping the alignment and competence of individuals, received insufficient attention.

There is evidence that leaders' personalities affect organizational performance and sustainability, especially in business management field, where the relationship has received much attention (Oreg & Berson, 2018; Birknerová & Uher, 2021). In order to better explore the positive consequences of leaders' personalities, the measurement of them becomes important. On the basis of personality scales for individuals in general, some scholars have improved and developed

the occupational personality scale fitting for party and government leaders, as well as launched some empirical studies. For example, Ao (2004) conducted a questionnaire survey on 745 governmental leaders in China using the MBTI personality scale, and came up with the distribution of personality types of Chinese leaders. Li (2006) and Yu (2019) respectively measured the personality of 183 civil servants in middle positions and 214 grassroots cadres in disciplinary inspection and supervision based on the Big Five Personality Inventory, revealing the impact of different dimensions of personality on the job performance of cadres. Xue (2017) used a sixteen-factor scale to measure the personalities of 127 division leaders and found that some personalities were significantly related to their leadership styles. Finally, Wang & Cui (2008) measured the personalities of 172 grassroots party and government leaders through a self-developed Chinese personality scale and the Big Five personality scale respectively, and found that the Big Five personality scale was worse than the Chinese scale in predicting job performance, thus suggesting the necessity of developing a personality scale for Chinese party and government leaders.

Due to the political and professional sensitivities, the use of personality scales to evaluate the personalities of party and government leaders presents notable constraints in scientific research and various applications. As the extant literature, current research primarily concentrates on the leaders at middle or grassroots levels, with limited inclusion of higher-ranking leaders. To some extent, it underscores the limitations inherent in personality scales and the complexities in appraising the personalities of senior party and government officials. Consequently, in the wake of the widespread adoption of data mining, many scholars posited that an alternative approach can be employed to assess the personalities of party and government leaders (Mairesse et al., 2007). In line with personality behavior theory, it is well-established that an individual's personalities exert a significant influence on their preferences, attitudes, language, decision-making, and behaviors (Mate, 2021). Therefore, we can deduce an individual's latent personalities by scrutinizing their external attitudes and behaviors. Consequently, by leveraging data mining and information intelligence analysis as technical tools for examining textual data stemming from the speeches, interviews, biographies, and news reports of party and government leaders, it becomes feasible to distill their linguistic and behavioral attributes, thus enabling the identification and extraction of their personalities (Yan & Yang, 2022).

A few scholars have developed information intelligence algorithms and models for measuring the leaders' personalities. For example, Post (2003) postulated a methodology wherein the political personality profiles of leaders can be constructed by scrutinizing their biographies. The approach allows for the discernment of the persistent and enduring personalities that facilitate an enhanced comprehension of their political conduct. Winter (2005) contributed to this discourse by introducing and validating an objective technique for remote identification and quantification of the personality and motivations of political leaders, including U.S. presidents. This was accomplished through the meticulous content analysis of a range of textual materials, such as speeches and interviews, utilizing a representative subset of their utterances. Similarly, Dyson (2008) employed a textual analysis method to code the responses to foreign policy matters articulated by 12 British prime ministers in the House of Commons from 1945 to 2008. His findings suggested that intelligent techniques can consistently distinguish and categorize the personalities of diverse leaders, further demonstrating their validity in expounding upon and predicting their political comportment. Nevertheless, limitations do exist in prior literature in this field, especially in the background of China.

In summary, the gaps in prior studies were twofold. First, the predominant methodology for measuring the personalities of government leaders was the use of psychological and behavioral scales, of which the validity is contingent upon the subjective cognitive ability of investigators as well as the respondents' perceptions of their professional and political sensitivities. Second, although there were some information intelligence algorithms to explore leaders' personalities, they still had the limitation of not being enough objective due to the complex source of textual data, for example, speeches, interviews, and biographies, which were mainly self-reported. Unlike previous studies, we believed that the explicit indicator of personality should be behavior. Hence, we proposed to reverse infer the personality traits of leaders from their participation in activities. The paper thus introduced the Latent Dirichlet Allocation (LDA) topic model, an unsupervised machine learning algorithm adept at representing high-dimensional text word spaces in the form of low-dimensional topic spaces. Its versatile applications extended across diverse domains, encompassing topic mining, text retrieval, text classification, and social network analysis (Zhang & Zhang, 2022). By integrating the LDA with factor analysis methods, the research endeavored to develop a novel algorithm tailored for the identification and analysis of personality traits. The algorithm operates through the clustering of reports on government activities authored by party and government leaders. The principal contribution of our study is situated at the theoretical level, augmenting existing research on the personality theory of party and government leaders. In practical terms, the findings herein have the potential to furnish valuable insights for the selection and management processes pertaining to prospective party and government leadership cadres.

## 2. Methodology

### 2.1 Research Step

For devising a method to analyze the personalities exhibited by party and government leaders, a structured research plan was formulated. The research entails the following steps.

Step 1: data collection and processing. Inception of the research journey involves the development of a Python-based web crawling program, dedicated to the acquisition of textual data pertaining to news reports concerning the political activities of leaders. The data is sourced from official websites. Subsequently, a series of data refinement procedures is executed, encompassing the removal of erroneous text segments, Chinese word segmentation, integration of a customized lexicon, and the filtration of disqualifying terms.

Step 2: thematic clustering of political activity reports. The meticulously curated textual data undergoes a process of thematic organization. The data are employed to construct LDA thematic models for each leader's political activity reports. The determination of the optimal number of topics is guided by the topic consistency parameter. This is succeeded by the visual representation of topics to unveil the structural underpinning of topic distribution, followed by an in-depth analysis of the interrelationships among keywords, leading to the nomenclature of identified topics. The topic intensity statistic is subsequently employed to quantify the significance of party and government leaders within various thematic domains. Ultimately, these topics, extracted from the political activity reports of all leaders, are amalgamated and succinctly summarized to form a comprehensive thematic database.

Step 3: personality analysis. The step encompasses the application of factor analysis to unearth common factors that encapsulate the overarching thematic information. Moreover, it involves a comparative evaluation of existing individual personality theories, aiming to elucidate theoretically-grounded interpretations and nomenclature for the common factors. Conclusively, it seeks to contribute to the development of a comprehensive theory pertaining to the occupational personalities of party and government leaders.

### 2.2 Participant (Subject) Characteristics

Appropriate identification of research participants is critical to the science and practice of psychology, particularly for generalizing the findings, making comparisons across replications, and using the evidence in research syntheses and secondary data analyses. If humans participated in the study, report the eligibility and exclusion criteria, including any restrictions based on demographic characteristics.

### 2.2 Methods

First, LDA topic model. The model is a Bayesian probabilistic model with a three-level hierarchy (document-topic-word) proposed by Blei et al. (2003). It can automatically encode a number of documents into a certain number of topics, based on the principle of identifying potential topic information in a large-scale document set or corpus through clustering. That is, it is a method to explore topics in a large number of texts and is well suited for document clustering, information retrieval and feature selection for unstructured text (Chen & Li, 2019).

The "document-topic-word" generation process of the model is as follows. Initially, the polynomial distribution  $\theta_m$ , representing the topic allocation for document  $m$ , is stochastically generated using the Dirichlet prior parameter  $\alpha$ . Subsequently, the topic  $Z_{m,n}$  for the  $n$ th word in document  $m$  is randomly drawn from the distribution  $\theta_m$ . Then, the polynomial distribution  $\psi$ , characterizing the word allocation for topic  $Z$ , is generated at random using the Dirichlet prior parameter  $\beta$ . The word  $W$  is then generated by jointly considering the topic  $Z$  and its associated word distribution  $\psi$ . The entire process is iterated  $n$  times. Dirichlet prior parameter  $\beta$  generates the polynomial distribution  $\psi_k$  of the words corresponding to the topic  $Z_{m,n}$ , and then combines the topic  $Z_{m,n}$  and its word distribution  $\psi_k$  to generate the words  $W_{m,n}$ . Finally, repeating for  $n$  times, a document  $m$  containing  $N_m$  words is generated, which generates a set of  $M$  documents under  $K$  topics. LDA gives the topic of each document in the set in the form of probability distribution, and by analyzing the whole document set and extracting its topic distribution, topic clustering can be realized based on the distribution.

The model training adopts the Gibbs sampling algorithm, and the model training needs to set in advance the Dirichlet distribution parameters  $\alpha$ ,  $\beta$ , and the optimal number of topics  $k$ .  $\alpha$  and  $\beta$  are generally selected as the default values, and  $k$  is determined by topic consistency in the study. The parameter refers to the semantic similarity of the high-frequency words appearing in each topic generated by the model, and a higher consistency score indicates higher model interpretability (Röder et al., 2015). Topic consistency is calculated by equation (1).

$$C_k = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m^k, v_l^k)+1}{D(v_l^k)} \quad (1)$$

Where  $V^k = (v_1, \dots, v_M^k)$  is the list of  $M$  most probable words in topic  $k$ ,  $D(v)$  is the number of documents containing word  $v$ , and  $D(v, v')$  is the number of documents in which words  $v$  and  $v'$  co-occur at least once. In the study, a document means a crawled news report on the political activities of party and government leaders.

In order to link the topic analysis with subsequent factor analysis and personality extraction, the distribution of topics needs to be numerically characterized and processed. The statistic chosen for the study is topic intensity. It refers to the relative weight of each topic in the corpus after clustering, reflecting the preference and importance of party and government leaders in attending different topic types. The higher topic intensity represents the more preference for the topic type of activity. The topic intensity is calculated by equation (2). Where  $P_k$  represents the intensity of the  $k^{\text{th}}$  topic word,  $N$  is the number of news report texts, and  $\theta_{ki}$  is the probability of the  $k^{\text{th}}$  topic in the  $i^{\text{th}}$  sample.

$$P_k = \frac{\sum_i \theta_{ki}}{N} \quad (2)$$

Second, factor analysis. It is a sophisticated multivariate statistical technique designed to extract latent common factors from the original data set comprising multidimensional indicators. It serves to streamline the structural complexity inherent in the initial data, thereby achieving the goal of dimensionality reduction (Yang, 2018). It explores a more nuanced representation of the research objects by encapsulating the majority of information from the original data set into a more concise set of comprehensive indicators. We leveraged the method to distill and amalgamate the topics encompassing news reports on the political activities of party and government leaders. The approach facilitates a refined analysis of personalities and the subsequent taxonomy of comprehensive thematic information vis-à-vis prior individual personality theories. The objective is to formulate a theoretical framework delineating the personalities exhibited by party and government leaders. Given the method as a mature algorithmic procedure, its computations may be readily performed through various statistical tools such as SPSS, obviating the need for a redundant explication of calculation steps and formulas in our study.

### 2.3 Data Collection and Processing

In order to validate the robustness of the model and method, we procured textual data concerning the political activities of party and government leaders in 31 provinces, municipalities, and autonomous regions of mainland China for the purpose of empirical analysis. The data collection commenced at the inception of the tenure of party and government leaders in the respective regions, extending through the culmination of their terms or up until December 2021. The data were collected from March to May 2022. The samples encompassed party and government leaders slated to serve a one-year term in the corresponding districts, or those who had served during the term preceding December 2021. The selection ensured an ample corpus of news data for comprehensive text analysis. Owing to the distinct spheres of responsibility held by deputy leaders, their participation in political activities typically correlates more strongly with the division of duties and responsibilities inherent in their roles, and less with individual personalities. As a result, we opted to focus our analysis on the principal party officials of each region, namely, party secretaries and governors.

The entirety of the data was acquired from the official websites of governmental authorities and the official online portals of established news media outlets within each region. For the retrieval of real-time news content featuring the names of party and government leaders, we employed a Python web crawling mechanism. Each instance of raw news data extracted through the process encompasses three essential fields, namely, the release time, the news title, and the news content. Subsequently, these data components were amalgamated to construct a comprehensive text database.

To process the textual data, the following steps were undertaken. First, data cleaning employing Excel tool. Given the inherent structural disparities in government website page layouts and the dynamic news plate diversification, the retrieval of original news data through full-text search was plagued by source-specific structural variations and a substantial degree of duplication. Consequently, the ensuing data cleansing process was necessitated. Primarily, it involved the elimination of news data characterized by high textual overlap in the main body, thereby retaining only the most comprehensive content. Subsequently, text content that proved empty or erroneous due to abnormal web crawling situations, such as the failure of website news page links or inaccessible news report videos, was systematically discarded. Furthermore, to mitigate errors stemming from full-text search results, irrelevant news data was filtered using heuristic methods, particularly by examining the first paragraph of the content. Second, Chinese text segmentation. The initial step involved the utilization of the Jieba thesaurus for text segmentation. To enhance the effectiveness of segmentation, the incorporation of custom dictionaries and deactivation word lists was imperative. Subsequently, the cleaned data was processed using the Pandas package, employing the deactivation word list from the Harbin Institute of Technology, China, while continuously augmenting this list with new deactivation terms derived

from subsequent word frequency analysis. The continuous refinement ensured the systematic removal of irrelevant characters from the Chinese text corpus (Yu & Zhao, 2018). Finally, to further refine the accuracy of word segmentation, a tailored dictionary was created. In the study, word frequency analysis was conducted on basis of text segmentation, and high-frequency terms were integrated into the user dictionary, aligning them with the content of the original text. Certain terms such as high-quality development, five-type government, great protection of the Yangtze River, and business environment were intentionally preserved in their original form, avoiding further dissection and alteration.

Following the processing step, a total of 56,804 valid texts were obtained. Within the dataset, it was observed that the mean number of news texts authored by regional party leaders stands at 903.42, with a standard deviation of 336.17. The maximum and minimum values within this category are 1,537 and 385, respectively. Similarly, the mean number of news texts attributed to government leaders averages 928.97, with a standard deviation of 329.36, and the maximum and minimum values in this category are 1,441 and 352, respectively. A statistical analysis revealed that there are no substantial differences in the number of news reports across the two categories.

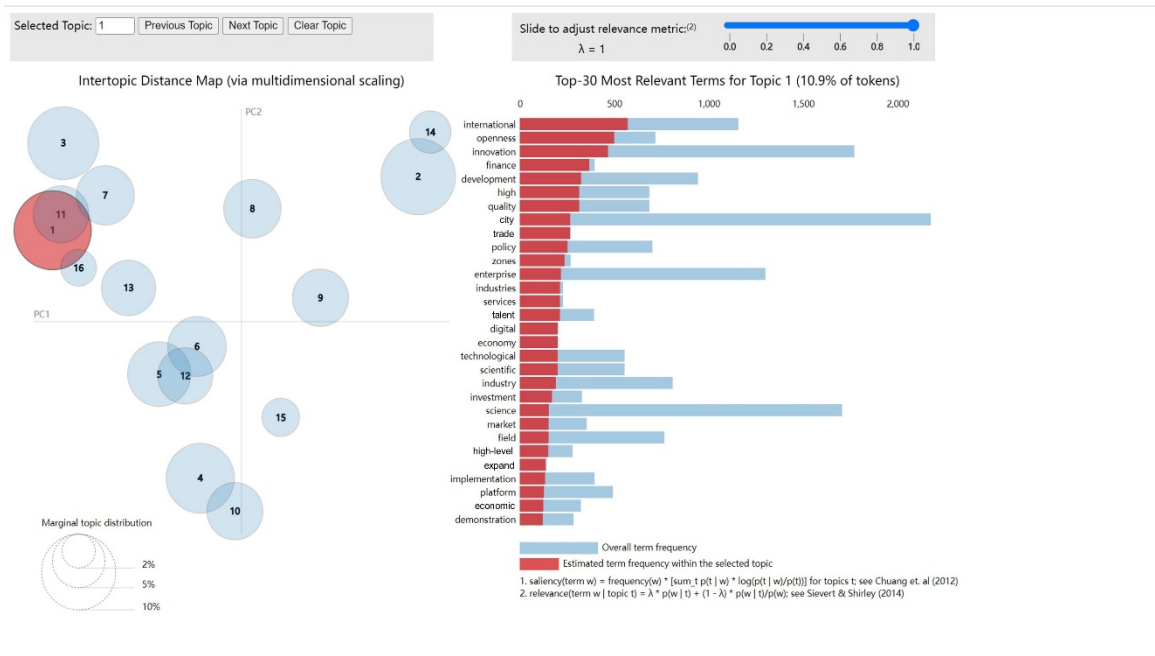
### 3. Results

#### 3.1 LDA Analysis

It involved the analysis of textual data encompassing news reports on the political endeavors of party and government leaders across 31 Chinese provinces, with the objective of constructing 62 distinct LDA topic models. A case study was undertaken, specifically focusing on the thematic extraction from news reports featuring the leaders of Beijing government. The case study elucidated the four key phases in LDA analysis, including ascertaining the optimal number of topics, visualizing topics, computing topic intensity, and extracting comprehensive topics.

First, determination of the optimal number of topics. The LDA model was generated utilizing the Python *gensim* library. Both hyperparameters  $\alpha$  and  $\beta$  in the model were configured to their default values of  $50/K$ , where  $K$  represented the number of topics determined based on topic consistency. Various approaches were employed to calculate topic consistency, such as  $c_v$ ,  $u\_mass$ ,  $c\_npmi$ , and  $c\_uci$ . In our study,  $c_v$  emerged as the most consistent metric in terms of interpretability and accuracy. The computation of  $c_v$  was illustrated in the equation (1). The number of topics was systematically traversed from 0 to 30, employing 50 iterations, with each topic model employing 20 keywords. Subsequently, each model underwent training, and its corresponding topic consistency score was calculated. The case sample exhibited maximum topic consistency at  $k=16$ , thus enabling the determination of the optimal number of topics for Beijing leaders, which was confirmed as 16.

Second, topic visualization. The topics derived from the LDA model were systematically visualized using the *pyLDAvis* package, a dedicated visualization tool. The outcomes of this visualization process were presented in Figure 1. In Figure 1 (a), each circle within the visualization represented an individual topic. The size of each circle corresponded to the volume of texts associated with the specific topic, while the spatial separation between circles served as an indicator of the degree of similarity between these topics. Figure 1 (b) focused on the topic 1 and offered insights into the 30 most pertinent terms linked to this topic. The word selection was performed to illuminate the central attributes of the topic. Furthermore, it is important to note that the visualization of various topics can be conveniently switched by either entering distinct serial numbers within the "selected topic" text box situated in the upper left corner of Figure 1 or by selecting different topic circles within the visualization.



(a) Topic generation and distribution (b) High frequency keywords for topic 1

Figure 1. Topic visualization of the sample case (Beijing)

Third, calculation of topic intensity. Based on the topic visualization results, the extracted topics were named by analyzing the high-frequency keywords of different topics and their commonalities. Further, the intensity of each topic was calculated according to the equation (2). For example, the case sample was extracted to 16 topics, and we carried out topic naming and topic intensity calculation based on their high-frequency keywords. The results were shown in Table 1.

Table 1. Topics of the sample case (Beijing) and their intensity and keywords

No.	Topics	Intensity	High frequency keywords
1	Review conference	0.037	system, governance, legislature, deliberations, justice, laws, reports, modernization, socialism
2	International cooperation	0.034	internationalization, airports, trade, education, economic zones, international engagement, leading groups, trade fairs, two zones, openness
3	Party history learning and education	0.072	communist party of China, socialism, party history learning and education, masses, celebration, nation, revolution, political consultative conference, NPC standing committee, career
4	Cultural preservation	0.053	culture, conservation, urban, research, parks, neighborhoods, inspection, old city, inscription, cultural relic
5	Industrial cooperation	0.046	cooperation, Beijing-Tianjin-Hebei synergistic development, field, technology, innovation, delegation, industry, region, Xiongan new area
6	Industrial innovation	0.129	enterprise, innovation, industry, technology, international, technology innovation, Zhongguancun, talent, platform, policy
7	Street governance	0.036	streets, points, complaints, rankings, districts, businesses, townships, ratings, groups, governance, topic parks

8	Support assistance and	0.023	poverty alleviation, rural revitalization, employment, support, agriculture, counterpart, countryside
9	Ecological governance	0.042	ecology, safety, risk, finance, green development, flood control, protection, governance, green, ecological protection
10	Epidemic prevention and control (society)	0.080	epidemic prevention and control, prevention and control, community, citizen, safety, health, mass, epidemic, epidemic prevention, public health
11	Epidemic prevention and control (organization)	0.109	prevention and control, leading group, epidemic prevention and control, epidemic, new crown pneumonia, epidemic risks, health joint controls
12	Winter Olympics	0.064	winter Olympics, organizing, OCOG, venues, Paralympics, tests safety, events, internationalization
13	Topic education	0.095	rectification, research, grassroots, party building, masses, governance, standing committee, topic education, cadres, receiving complaints, governance party
14	Community governance	0.040	community, garbage, masses, governance, streets, residents neighborhoods, property management, life, citizen
15	Livelihood governance	0.073	cities, research, decongestion, education, engineering, high-quality development, systems, schools, action plans, students, governance
16	National Congress report	People's (NPC) 0.070	review conference, reports, NPC, 14 <sup>th</sup> five-year plan, NPC standing committee, people's livelihood, draft, economic and social development, municipal people's political consultative conference

Fourth, summary and integration of topics. Upon summarizing the thematic content and keyword distributions related to 62 party and government leaders, a notable observation emerges, i.e., there exists a pronounced semantic convergence and similarity among them. Even the thematic underpinnings of political activities undertaken by the same leader tend to exhibit substantial resemblances. For example, the topic 10 "epidemic prevention and control (society)" and topic 11 "epidemic prevention and control (organization)" manifested a significant correlation and were integrated to a topic "epidemic prevention and control" in terms of the similar high-frequency keywords. Furthermore, an exploration of the original news texts uncovered a multitude of topics within the realm of governmental gatherings that naturally segregate into various sub-thematic categories including NPC meetings, deliberative assemblies, and thematic educational initiatives. In response, we harnessed the insights obtained from topic visualization outcomes and coalesced them with the content extracted from the original news. The synergy enabled us to unite sub-topics displaying marked semantic content overlap. Consequently, we successfully distilled all political activities of party and government leaders into 17 overarching thematic categories including industrial collaboration, international cooperation, industrial innovation, technological and talent innovation, business environment, industrial research, project research, poverty alleviation and rural rejuvenation, support and assistance initiatives, epidemic prevention and control and resumption of work and production, mass security and production control, tourism and cultural promotion, grassroots visits and condolences, livelihood governance, social security governance, government affairs meetings, and ecological governance.

### 3.2 Factor Analysis

Factor analysis was employed as a methodological approach to scrutinize the thematic intensity exhibited by 62 party and government leaders. They actively engaged in 17 distinct thematic categories of activities. It was conducted by using SPSS. Six principal factors were extracted, with the eigenvalues surpassing the threshold of 1. The detailed variance contributions and their cumulative effects were outlined in Table 2. The cumulative variance contribution rate reached an impressive 65.360%. It underscored the robustness of the extracted factors in encapsulating the diverse topics permeating the activities of party and government leaders under investigation.

Table 2. Total variance explained

Factors	Initial eigenvalue			Extract the sum of the squared loads			Rotational sum of squared loads		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	2.528	14.871	14.871	2.528	14.871	14.871	2.089	12.287	12.287
2	2.094	12.316	27.188	2.094	12.316	27.188	1.944	11.437	23.724
3	2.012	11.833	39.021	2.012	11.833	39.021	1.886	11.093	34.817
4	1.908	11.224	50.244	1.908	11.224	50.244	1.863	10.961	45.779
5	1.503	8.842	59.086	1.503	8.842	59.086	1.807	10.630	56.408
6	1.067	6.274	65.360	1.067	6.274	65.360	1.522	8.952	65.360

We employed orthogonal rotation to the factor loading matrix by maximizing variance. The rotation process achieved convergence after 10 iterative steps. To facilitate a clear interpretation of each factor, we presented the rotated factor loading in Table 3, ordered by the magnitude of their coefficients.

Table 3. Rotated factor loading matrix

Topics	Factors					
	F1	F2	F3	F4	F5	F6
Livelihood governance	<b>0.834</b>	-0.020	-0.180	-0.023	-0.081	-0.122
Project research	<b>0.641</b>	0.156	0.095	0.239	0.243	0.398
Business environment	<b>-0.467</b>	0.167	-0.045	0.186	0.031	-0.339
Government affairs meetings	-0.518	<b>-0.721</b>	-0.116	-0.035	0.004	0.138
Industry research	-0.096	<b>0.648</b>	0.117	-0.215	-0.090	0.178
Mass security and production control	-0.301	<b>0.637</b>	-0.055	0.043	0.040	-0.098
Social security governance	-0.104	<b>-0.626</b>	0.207	-0.076	0.060	0.111
Ecological governance	0.194	-0.031	<b>-0.794</b>	0.061	-0.104	-0.060
Epidemic prevention and control, and resumption of work and production	0.461	-0.177	<b>0.704</b>	0.119	-0.015	-0.024
Grassroots visits and condolences	0.103	0.061	<b>-0.647</b>	-0.087	0.329	0.439
Support and assistance initiatives	-0.063	0.041	0.086	<b>0.846</b>	-0.101	0.023
Poverty eradication and rural revitalization	-0.226	0.108	-0.091	<b>-0.730</b>	0.073	0.053
Culture and tourism promotion	0.247	0.009	0.244	<b>-0.474</b>	0.047	0.067
International cooperation	0.031	0.103	0.005	0.113	<b>-0.916</b>	-0.041
Industrial cooperation	0.032	-0.009	-0.024	-0.149	<b>0.811</b>	-0.429
Technological and talent innovation	-0.026	0.247	0.213	0.167	0.237	<b>-0.734</b>
Industrial innovation	0.069	0.184	0.322	0.393	-0.190	<b>0.481</b>

Factor F1 exhibits substantial loading on three thematic categories, i.e., livelihood governance, project research, and business environment. Upon a comprehensive examination of the original news reports, it becomes evident that these topics are manifested through the active involvement of party leaders in livelihood-related activities, encompassing education, healthcare, and employment initiatives. Additionally, they engage in research activities, notably in the initiation and execution of large-scale projects, and also contributing to the enhancement of the local business environment. This pattern of behavior suggests a propensity among the leaders for rational and pragmatic approaches. They demonstrate a willingness to scrutinize, deliberate, and resolve practical issues.



Factor F2 exhibits notable loading on four topics, i.e., government affairs meetings, industrial research, mass security and production control, and social security governance. The behaviors associated with this factor are characterized by the active participation of party and government leaders in presiding over or attending a multitude of meetings addressing various government and party affairs. Furthermore, they exhibit a proclivity for conducting visits and research activities. They place significant emphasis on addressing daily but intricate issues for preserving social order and stability. The issues include endeavors such as disaster response, safety protocols, and the eradication of corruption and malpractice.

Factor F3 exhibits notably elevated factor loading across three thematic domains, i.e., ecological governance, epidemic prevention and control and resumption of work and production, and grassroots visits and condolences. Leaders who excel in these domains manifest a proclivity for hosting events related to ecological governance, encompassing activities that pertain to environmental preservation and sustainable development. Furthermore, they frequently take the initiative in front-line guidance for activities concerning epidemic prevention and control, medical assistance, resumption of work and production, in addition to extending their presence to visit and offer solace to front-line workers and retired cadres. The execution of these thematic activities necessitates a keen awareness of prevailing social conditions and a proficiency in identifying and addressing issues through direct engagement with grassroots constituents. Consequently, the factor underscores the leaders' commitment to societal responsibilities and ethical imperatives.

Factor F4 registers substantial factor loading across three principal thematic spheres, i.e., support and assistance initiatives, poverty alleviation and rural revitalization, and culture and tourism promotion. A considerable cohort of party and government leaders affiliated with these thematic domains actively engage in and execute support and industrial assistance initiatives, as well as exhibit an unwavering commitment to the vigorous advancement of poverty alleviation and rural rejuvenation. Simultaneously, they ardently foster the local tourism sector while intensifying efforts in promoting indigenous cultural traits and characteristics. The aforementioned thematic activities are instrumental in fostering local economic and social development, and concurrently, they generate highly favorable word-of-mouth effects. Significantly, the implementation of these thematic activities directly contributes to the enhancement of performance metrics that have garnered heightened scrutiny in recent years, thereby underscoring the pragmatic and realistic attributes inherent to party and government leaders.

Factor F5 within the public domain exhibits a pronounced affinity for topics centered around international cooperation and industrial collaboration. A comprehensive analysis of original news reports indicates that these topics are intricately linked with governmental endeavors, including international cooperative initiatives aimed at facilitating global engagement (e.g., exemplified by One Belt, One Road), as well as interregional industrial cooperation efforts spanning areas such as Beijing, Tianjin, Hebei, Guangdong, Hong Kong, Macao, the Pan-Pearl River Delta, and the Chengdu-Chongqing region. Factor F5 serves as a testament to the proclivity of party and government leaders towards openness and international engagement.

Factor F6 is characterized by its salient focus on topics pertaining to technological and talent innovation, and industrial innovation. The behavioral manifestations of this factor predominantly manifest in the ardent support and emphasis exhibited by leaders towards collaborations involving industry, academia, and research institutions, along with their proactive orchestration of initiatives in the realm of science and technology innovation. Furthermore, it is marked by the adoption of policies that actively bolster industrial innovation. The factor underscores the proclivity and attentiveness of party and government leaders towards fostering a culture of innovation and technological advancement.

We conducted an extensive literature review, encompassing numerous theories concerning individual and occupational personality, including Big Five personality framework, the Myers-Briggs Type Indicator (MBTI), the Nine Personality Types theory, etc. (Phan & Rauthmann, 2021). Ultimately, our investigation revealed a congruence between Hollander's occupational personality theory and the outcomes derived from the aforementioned factor analysis. Hollander's theory categorizes individual personalities linked to vocational interests into six types, i.e., realistic, research-oriented, artistic, social, entrepreneurial, and conventional (Holland, 1997). Each type is characterized by distinct external behavioral attributes. Individuals exhibiting a high inclination towards the realistic personality are inclined towards hands-on implementation, modesty, and independent work, yet may display limitations in social interaction due to a conservative disposition. Those scoring high on the research personality demonstrate proficiency in rational and abstract thinking, along with a tendency to accept and engage in independent tasks. However, their relatively weaker social skills and hesitance towards action can be perceived as limitations. The artistic personality is marked by creativity and a penchant for aesthetics, emphasizing innovation but sometimes falling into the trap of idealism at the expense of practicality. Social personality prioritizes engagement with social issues and exhibits

proficiency in interpersonal interactions. The entrepreneurial personality is characterized by high openness, a proclivity for competition and adventure, and a readiness to initiate new ventures, although an inclination towards power-seeking and a driven approach may be observed. Individuals with a strong conventional personality preference excel in adhering to structured plans and systems and pay meticulous attention to detail. However, their cautious and conservative tendencies may result in a lack of innovation.

Subsequently, we correlated the six factors extracted from factor analysis with the six personality types established in Hollander's theory, denoting these factors as F1- investigative, F2-conventional, F3-social, F4-realistic, F5-enterprising, and F6-artistic. We then assigned them to the six dimensions encompassing the personalities of provincial party and government leaders. To sum up, we provided a comprehensive synthesis and depicted an analytical framework illustrating personality theory and action preference among provincial-level party and government leaders in China, as shown in Figure 2.

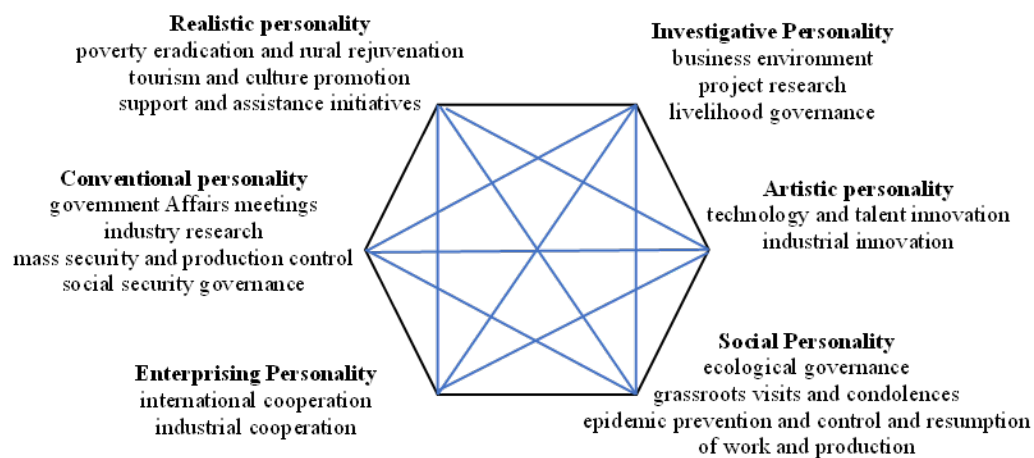


Figure 2. Personality theory and action preferences of party and government leaders

### 3.3 Spatial Analysis

We utilized factor analysis to compute scores corresponding to the six types of personality for the 62 party and government leaders, employing component score coefficients. Personality scores below zero were considered less prominent and were consequently standardized to zero. Subsequently, the remaining data were categorized based on inherent discontinuity points. To visualize the data, ArcMap software was employed to generate distribution maps depicting the personalities of the leaders, as presented in Figure 3.

Within the map, the color depth serves as a representative indicator of the scores assigned to personalities, with darker colors denoting higher score of a specific personality. The left column of the map illustrates the distribution of personality among party leaders, while the right column portrays the distribution of personality among government leaders. The differentiation in presentation facilitates comparative analysis.

Analysis of spatial aggregation of personalities among party and government leaders. As delineated by Ao (2004), discernible differentiation exists in the personality types exhibited by leading cadres across distinct regions. The variations may be intricately linked to the distinctive working environments and professional attributes characterizing leading cadres in diverse groups. As depicted in the Figure 3, a discernible spatial aggregation effect characterizes the personalities of party and government leaders. The spatial clustering phenomenon implicitly aligns with the established patterns of geographical disparities in China's economic and social development. Government leaders in the less developed northeast and western regions prominently exhibit investigative personality (Figure 3b), while their counterparts in these regions tend to manifest strong enterprising personality (Figure 3c). In contrast, the leaders along the prosperous east coast exhibit distinctively prominent social and realistic personalities (Figures 3e-h), whereas the geographic distributions of conventional and artistic personalities display less conspicuous regularity (Figure 3i-l). The observed clustering effect, along with its associated regularities, underscores the potential predictive utility of the personalities among party and government leaders, including social, realistic, investigative, and enterprising

personalities, in elucidating the economic development levels within regions and their inter-regional disparities. Furthermore, the prevalence of social and realistic personalities among party and government leaders appears to enhance their adaptability to the provincial and local conditions prevalent in the developed eastern regions.

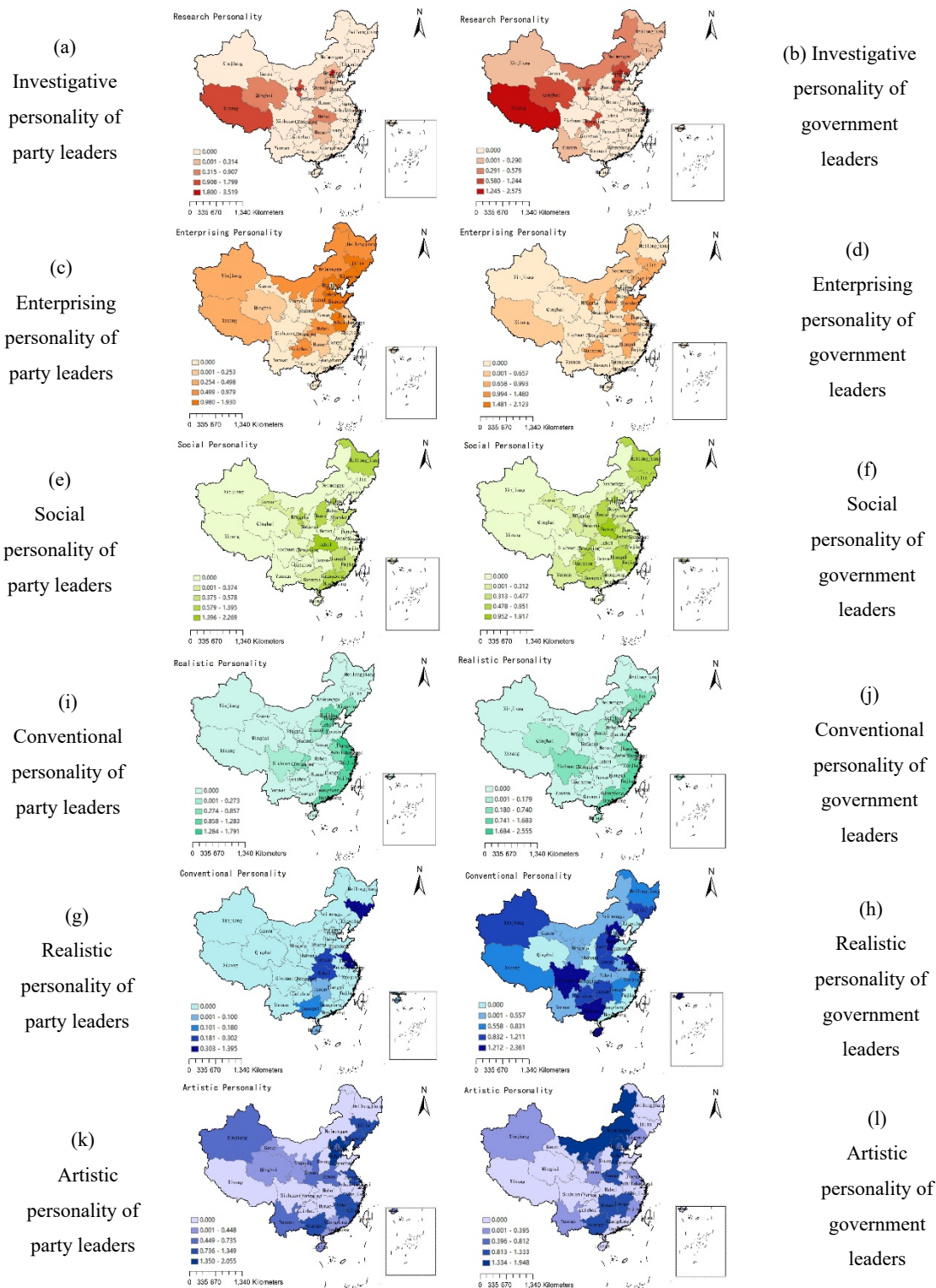


Figure 3. Spatial distribution and aggregation of personalities of party and government leaders in China

Analysis of inter-regional economic development disparities. The underlying causes of disparities in economic development between regions are multifaceted and have deep historical roots. The party leaders under examination in our study held office during their most recent terms, thereby raising the question of whether the personalities of these party leaders exert influence on local economic development or whether, conversely, local economic conditions and levels of development influence the outward expression of their professional personas. The spatial distribution of conventional and artistic personalities, as illustrated in Figure 3i-l, lacks a discernible regular pattern. It suggests that these particular personalities may not hold significant explanatory power with regard to the variations in regional development. It can be inferred from these findings that effective party and government leaders are characterized by their capacity to refrain from immersing themselves in mundane and routine transactional activities, and instead, they must exercise prudence in avoiding unrealistic notions and speculative endeavors.

Exploring Hollander's theory of occupational personality. Hollander's theory of occupational personality posits the existence of three sets of relative relationships among six personalities: investigative versus enterprising, social versus realistic, and conventional versus artistic. There exists minimal common ground between personalities within these relative relationships, and only a few individuals manage to maintain elevated levels across both traits. In contrast to the depiction in Figure 3(i-d), an examination of party and government leaders reveals a substantial conformity to this pattern in their expression of investigative and enterprising personalities. In contrast to the leaders operating at the grassroots level, senior leaders exhibit a reduced extent of personal exposure to comprehend and engage with the realities of grassroots contexts. Consequently, their proclivity towards an investigative personality is not as conspicuous. Owing to the augmented decision-making authority vested in senior leaders, their enterprising personality assumes a more salient profile compared to that of grassroots leaders (Sun et al., 2011). The distributions of these two traits exhibit minimal overlap, and intriguingly, there appears to be potential complementarity between the investigative personality of government leaders and the enterprising personality exhibited by party leaders. However, when scrutinizing the distribution of social and realistic personalities (Figure 3e-h), a deviation from the established pattern becomes apparent. Both party and government leaders in the eastern seaboard regions display elevated levels of both social and realistic personalities.

Examining leadership in economically developed regions. It is reasonable to assert that governance in economically advanced regions poses a substantial challenge, demanding a comprehensive evaluation of both the pragmatic acumen and adeptness in communication and coordination among party and government leaders. Consequently, the study casts fresh illumination upon the necessity for capable party leaders to exhibit a synthesis of intricate yet moderated personalities. Furthermore, they must adeptly harmonize ostensibly contrasting yet mutually reinforcing personalities, maintaining them at elevated levels and in equilibrium. In essence, proficient leaders exhibit a pronounced proficiency in the adroit management and harmonization of practical affairs and social relationships, executed with a high degree of efficiency.

## 4. Conclusions

### 4.1 Findings

Research in the domain of personality computing is already in progress, as evidenced by previous work (e.g., Phan & Rauthmann, 2021). This field explores the effective integration of contemporary personality theory and computer science to facilitate practical personality assessments. Conventional questionnaire-based personality tests prove inadequate for addressing certain challenges, such as the assessment of political leaders, cybercriminals, and other sensitive cohorts, as well as the demands of big data-driven business intelligence focusing on consumer personality analysis. Hence, personality computing holds substantial promise for practical applications.

In the scope of this study, we undertake the construction of a LDA topic model and integrate factor analysis technique to formulate an algorithm conducive to the recognition and analysis of personalities. Subsequently, we successfully apply this algorithm within the context of personality analysis among party and government leaders in the Chinese context. Drawing upon the empirical findings of our research, we are positioned to deduce the following conclusions. First, the utilization of the LDA topic model proves to be highly suitable for the extraction and clustering of thematic content within news reports pertaining to the political activities of party and government leaders. Additionally, factor analysis serves as a valuable tool for disentangling and identifying the overarching topics while deciphering the dimensions of personality influenced by public factors. The synergistic amalgamation of these two analytical methods results in the successful achievement of personality recognition and analysis among party and government leaders. Second, our empirical analysis demonstrates that the personalities of party and government leaders encompass six categories, i.e., realistic, investigative, artistic, social, enterprising, and conventional dimensions. These categories align closely with Hollander's theory of occupational personality. Third, spatial analysis unveils a noteworthy spatial

distribution pattern within the personalities of party and government leaders. This pattern not only aligns with our cognitive expectations but also possesses the capacity for explanation and prediction. Moreover, it yields insights of greater managerial significance when compared to Hollander's theory of occupational personality. In summary, this study accomplishes its objectives by introducing a method for the systematic analysis of the personalities of party and government leaders. Furthermore, it substantiates the scientific validity and effectiveness of this method.

In fact, Post (2010) in the book *The Psychological Assessment of Political Leaders: With Profiles of Saddam Hussein and Bill Clinton* discussed in detail the ideas and methods for evaluating and measuring the personalities of political leaders. Post believed that there are two schools of attention to the leader's personality: one focuses on the objective fact and composition of the leader's personality, while the other focuses on the action mode of the personality, including the influence of the working situation on the leader's personality and the relationship between the leader's social roles and the expected explicit personalities. The former is usually measured by questionnaire survey, while the latter is measured according to the logic of personality-leadership-behavior, so the leader's personality can be evaluated by observing the leader's behaviors. Our study is actually practicing the view of the second school of thought. In addition, Winter (2005) proposed that motivation and personality can be measured indirectly and objectively. He believed that personality externalization is embedded in the social situations. In a political situation, since the behavior of a political leader can be predicted by his or her personality, we can infer the personalities of the leader through his or her behavior. However, in the study of Winter (2005), leaders' personalities were interpreted by the Big Five personality theory, while our study was supported by Holland's occupational personality theory, which may be more suitable for interpreting occupational situations. Compared with the findings of Winter (2005), our study has made a contribution to expand the theory. Finally, different from policy response text data used in Dyson's(2008), our study used the daily behavior data of Chinese government leaders for analysis. Therefore, the conclusions of our study may be more in line with the objective facts of the leader's personalities.

#### 4.2 Implications

The contributions of this study are twofold. First, we have rigorously integrated and developed algorithms and models tailored to the analysis of personality exhibited by party and government leaders. The methods possess the distinct advantage of operating remotely, thereby not necessitating direct interaction with the leaders in question. Furthermore, they maintain objectivity by eschewing reliance on self-reports provided by these individuals. Second, we have introduced a theoretical framework for characterizing the personalities and action preferences of party and government leaders. Our investigation has substantiated the framework's explanatory power, particularly in terms of its ability to elucidate spatial heterogeneity in regional development. The empirical validation opens the door to further research endeavors, promising to shed light on the far-reaching consequences of leaders' personalities in the future.

In light of the study's findings, we posit three strategic management recommendations. First, the algorithms and models developed herein can be deployed to establish a comprehensive personality database of party and government leaders, as well as other individuals whose digital footprints abound online (e.g., celebrities, online consumers). Such a database holds immense potential for enhancing applications in the domains of public management, social governance, and commercial enterprises. Second, the refined personality theory and action preference analysis framework put forth in this study can serve as guiding principles for the cultivation, selection, appointment, and evaluation of local party and government leaders. The conceptual framework stands to evolve into a pivotal tool in human resource management, specifically catering to the unique requirements of leaders. Third, in the process of appointing party and government leaders, it is imperative to meticulously consider the candidates' personalities and their intricate interpersonal dynamics. This necessitates an exploration of the complementary relationships between the personalities exhibited by the leaders. Moreover, an in-depth analysis of the interplay between provincial and local contextual factors and the personalities of these leaders is crucial. This multifaceted examination, elucidating the nexus between provincial and local conditions and the personalities of party and government leaders, is pivotal for informed and effective decision-making processes.

#### 4.3 Limitations and Future Research

The findings of this study do possess certain limitations with regard to their practical applicability.

First, the study focused exclusively on the top leaders of Chinese provincial party and governments, resulting in a small sample size, which limited the generalizability of our findings. Additionally, the clustering analysis applied to the news texts during the whole employment term of the leaders lacked temporal granularity, as the texts were not processed by year. Therefore, the cross-sectional model employed in our study lacked the capacity for dynamic analysis over time. To address the limitations, several adjustments can be considered. Primarily, we will broaden the scope of sampling in the future to the entire leadership team of the party and government organizations, in order to capture a

more comprehensive representation of the leadership. Furthermore, a commitment to conduct dynamic analyses of personality traits over the tenure time of the leaders is articulated. The temporal perspective is crucial for understanding how the traits evolve and manifest over the course of leadership responsibilities. Furthermore, an emphasis will be placed on exploring the complementarity and differences among the members of leadership teams. The multifaceted analysis intends to contribute to a nuanced understanding of the professional personality traits specific to government leadership. By incorporating these adjustments, the theoretical framework underpinning government leadership professional personality traits can be refined and expanded, fostering a more comprehensive understanding of the complexities inherent in this domain.

Second, in practice, the efficiency of the model proposed by our study deserves careful consideration. Particularly, inefficiencies were notably experienced during the process of crawling news reports and conducting data analyses pertaining to the political activities of party and government leaders across 31 regions. It underscores the necessity for further enhancement and optimization of the program and algorithm employed in these activities. In addition, in evaluating the results and conclusions, several limitations become apparent. Achieving an optimal alignment between the six common factors and the corresponding dimensions within Holland's occupational personality theory may necessitate more evidences to substantiate the findings. The extraction of six common factors through factor analysis yielded a total variance contribution rate that was relatively modest. This prompts consideration regarding the potential misalignment between the data obtained through thematic analysis and the methodologies applied in factor analysis. In future research, we will optimize the algorithms. For instance, the adoption of the LDA2vec model, coupled with the utilization of context vectors for predictive purposes, offer a promising avenue for improving the accuracy of topic recognition within a corpus containing intricate and extensive textual data.

Finally, when considering the practical application of the research methods, the widespread or commercial implementation of text crawling from governmental websites raises concerns about potential legal and ethical implications. Such consideration might impede the application of the methods and models proposed in our study. In addition, while the LDA topic model and factor analysis methods employed benefit from existing software tools to facilitate their execution, there remains a necessity for further exploration. The exploration should aim to seamlessly integrate the algorithms from diverse tools, enhancing the synergy and compatibility among them. Looking ahead, the trajectory of advancement encompasses not solely the refinement of theoretical constructs and algorithms but also a concerted effort toward devising innovative solutions capable of overcoming various barriers to application. Such proactive measures are pivotal for augmenting both the theoretical underpinnings and practical applicability of the study.

## References

- Ao, X. (2004). Studies of personality types of leadership in China. *Journal of Psychological Science*, 27(3), 731-734. <https://doi.org/10.16719/j.cnki.1671-6981.2004.03.058>
- Birknerová, Z., & Uher, I. (2021). Assessment of Management Competencies According to Coherence with Managers' Personalities. *Sustainability*, 14(1), 170. <https://doi.org/10.3390/su14010170>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(3), 993-1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Chen, J., & Li, Y. (2019). Social media fatigue research based on LDA topic model: Take WeChat as an example. *Information Science*, 37(12): 78-86. <https://doi.org/10.13833/j.issn.1007-7634.2019.12.012>
- Dyson, S. B. (2008). Text annotation and the cognitive architecture of political leaders: British prime ministers from 1945–2008. *Journal of Information Technology & Politics*, 5(1), 7-18. <https://doi.org/10.1080/19331680802149624>
- He, Q. (2020). Research on the credibility of cadre selection and appointment under the perspective of competency. *Research of Administration Science*, 62(2), 51-57.
- Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments. *Psychological Assessment Resources*.
- Ji, N., & Zhang, J. (2022). Cultivation promotion: An explanatory framework for the promotion of cadres. *Administrative Tribune*, 28(1), 100-110. <https://doi.org/10.16637/j.cnki.23-1360/d.2022.01.010>
- Li, N. (2006). *Research on relationship between the personality and job Performance of middle-level public servants*. Dalian, China: Dalian University of Technology.

- Liu, Y., Liu, M. T., & Shang, R. X. (2022). Four-dimension fit drives the modernization of national governance capacity: Case analysis of Chinese communist party's cadre selection practices. *Journal of Public Management, 19*(3), 35-45. <https://doi.org/10.16149/j.cnki.23-1523.20220629.001>
- Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research, 30*(1), 457-500. <https://doi.org/10.1613/jair.2349>
- Mate, M. (2021). Party leaders in Eastern Europe. Personality, behavior and consequences. *Europe-Asia Studies, 73*(7), 1391-1393. <https://doi.org/10.1080/09668136.2021.1951015>
- Oreg, S., & Berson, Y. (2018). The impact of top leaders' personalities: The processes through which organizations become reflections of their leaders. *Current Directions in Psychological Science, 27*(4), 241-248. <https://doi.org/10.1177/0963721417748397>
- Phan, L. V., & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and personality psychology compass, 15*(7), e12624. <https://doi.org/10.1111/spc3.12624>
- Post, J. M. (2003). Assessing leaders at a distance: The political personality profile. *The psychological assessment of political leaders: With profiles of Saddam Hussein and Bill Clinton*, 69-104.
- Post, J. M. (2010). *The Psychological Assessment of Political Leaders: With Profiles of Saddam Hussein and Bill Clinton*. Ann Arbor: University of Michigan Press.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).
- Shang, H. (2018). Paths to motivations and accountabilities: A review of China's 40-year government performance evaluation reform. *Chinese Public Administration, 8*, 85-92.
- Sun, Y.G., Niu, R., & Luo, X. (2011). The Research on Status quo of the Leaders' Personality Traits and Types of Administration Ability. *Journal of Dalian University, 32*(6), 97-101.
- Wang, D., & Cui, H. (2008). Predicting job performance of Chinese local government executives with QZPS and NEO PI-R. *Acta Psychologica Sinica, 7*, 828-838.
- Winter, D. G. (2005). Things I've learned about personality from studying political leaders at a distance 1. *Journal of personality, 73*(3), 557-584. <https://doi.org/10.1111/j.1467-6494.2005.00321.x>
- Xue, H. (2017). *Research on relationships between personality traits and leading styles of leaders on department level and intervention of social work*. Nanjing, China: Nanjing University.
- Yan, Y., & Yang, Z. (2022). Failure of policy attention allocation of local cadres and its governance. *Academia Bimestris, 5*(7), 128-134. <https://doi.org/10.16091/j.cnki.cn32-1308/c.2022.05.007>
- Yang, Y. (2018). Research on influence evaluation of Weibo topic communication based on factor analysis. *Information Science, 36*(7), 62-65. <https://doi.org/10.13833/j.issn.1007-7634.2018.07.011>
- Yu, W. (2019). *The effect of Big Five personality on the work performance of grassroots discipline inspection and supervision cadres: The role of psychological capital and self-monitoring in between*. Hangzhou, China: Zhejiang University.
- Yu, Y., & Zhao, N. (2018). Automatic selection of domain-specific stopwords in topic model of patent text. *Library and Information Service, 62*(11), 120-126. <https://doi.org/10.13266/j.issn.0252-3116.2018.11.014>
- Zhang, D., & Zhang, M. (2022). A review on application studies of LDA topic models in library and information science field. *Documentation, Information & Knowledge, 39*(6), 143-157. <https://doi.org/10.13366/j.dik.2022.06.143>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).