9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

One Image, Two Mirrors: Mapping Chinese Populist Ultranationalism on Weibo and Twitter

Keywords: Populism, Ultra-nationalism, Natural Language Processing, Weibo, Twitter

Extended Abstract

Introduction. Populism is widely accepted as an ideology that divides society into two homogeneous and antagonistic groups, "the pure people" versus "the corrupt elite" (Mudde, 2004). Even though populism has received much scholarly attention, it is less known whether populism can exist in a non-democratic context. Interestingly, China has seen a recent increase in animosity directed at foreigners, Westerners, and pro-Western Chinese intellectuals and political elites in the name of "the people" (Mattingly & Yao, 2020), which suggests a right-wing populist tendency. In recent years, as cyberspace became the main field that populism unfolds, massive online data and advanced computational techniques have provided scholars new possibilities to summarize both the consistency and differences of populism globally. This study creates a new approach to systematically categorize and analyze Chinese populist discourses by using computational approaches. The paper focuses on three research questions: How do **(RQ1)** topics, **(RQ2)** semantic content, and **(RQ3)** the main agenda setters and audiences of Chinese populist ultranationalism differ on Sina Weibo and Twitter?

Data and Methods. Data from Weibo and Twitter was collected in two rounds: First, there was an initial Weibo collection combing recent trending ("hot") events that are related to discussions around populism or reflect populist rhetoric. We coded events into populist and non-populist events, and we generated a keyword list about populism by using the event-based dataset. A second round of data collection was conducted using the keyword list to search for populism related content on Weibo and Twitter (date ranges from 1st May 2022 to 31st May 2022), yielding 241,920 Weibo posts and 560,770 Twitter posts, from which 50,000 posts from each platform were randomly selected for the purpose of analysis due to limitations of computational power. From these two sets, 10% of the posts were selected for manually labelling into either populist or non-populist content for further automated text classification. We also sought to identify the characteristics of users of the platforms. First, we classified account types on Weibo as journalists, government, professionals, we-media[1], and individuals. On Twitter, accounts are only classified as journalists, we-media, and individuals. Then posts were labelled into populist or non-populist. The 20,000 labeled posts were then, in a final step, analyzed with classifiers for training to be able to classify the remaining data.

Model Selection and Post-Analysis. The language model used was Roberta, a variant of BERT (Bidirectional Encoder Representations from Transformers) that is recognized as state-of-the-art and was trained on very large language corpora. In 2019, a Chinese developer published the first Chinese RoBERTa pre-trained model, which was trained on a 10 gigabyte (base) and 30 gigabyte (large) Chinese corpus including News, Question and Answer, and Encyclopedia data[2]. It outperformed other BERT models in most performance tests. This paper applied this model to our populism datasets. Having received the classification results, Latent Dirichlet

[1]Registered accounts with high numbers of followers.
[2]Github Source: https://github.com/brightmart/roberta_zh.

allocation (LDA) was used to measure the potential topics covered in the populist posts as it has powerful ability to identify hidden patterns from a large corpus (Blei & Lafferty, 2006). To investigate whether some populist posts reveal latent semantic meanings such as emotions, values, and ideologies, the correlations between platforms, targeted groups, and topics in low-dimension space were checked using Multiple Correspondence Analysis (MCA).

Results. 1. The performance of the Roberta has an accuracy of 86.0% on Weibo data and 92% on Twitter data and compares with the performance of base Bert model and the Fasttext model. It is somewhat higher than base Bert but both Bert family models clearly outperform Fasttext. Labels were obtained for 100,000 posts in the full data sets with 14.66% populist posts in the Weibo set and 8.43% populist posts in the Twitter set. This allows comparison of the proportion of populist content in each category on Weibo and Twitter (see **Figure 1**). **2. We also visualize topic distributions of populist posts** on Weibo and Twitter. N-gram based LDA is used to extract and has been able to capture the discrepancy of themes among populism categories and platforms. The name for each topic is assigned by selecting common words plus specific examples (**Figure 2**). **3. We identified the styles of semantic content of populism** using the low-dimensional Euclidean space. Because all we have are categorical variables, an MCA was used for correlations (see **Figure 3**). The variation in the variables is explained by 13.3% in the first dimension and 12.2% in the second. The first dimension and the second dimension can be seen on a Destructive-Constructive scale and the Populist-Ultranationalist scale, respectively. Overall, the four quadrants correspond to the following four semantic styles, they are: constructive and populist (Q1), destructive and populist (Q2), destructive and ultranationalist (Q3), and constructive and ultranationalist (Q4). **4. We also compared the types of users** who generated populist and non-populist content. **Figure 4** shows distribution of different types of accounts on both platforms. The links between agenda setters, information receivers [3], and targets are further illustrated in **Figure 5**.

Conclusion. This paper has compared the topics, semantic content, and protagonists of populist discourses on Weibo and Twitter. Using machine learning method to identify populist content, we find that populist ultranationalism is reflected in micro-blog posts on both Chinese domestic social media and international social media. Meanwhile, LDA classification of different topics behind populist content suggests five issues discussed on Chinese social media that are all related to populism: popular sovereignty, anti-foreigner, anti-western, anti-intellectual, and anti-elitism. Therefore, we have seen the bottom-up nature of populism in the sense that elites are targeted. This paper contributes to interdisciplinary studies by providing a deeper understanding of the integration of social science and computer science. The development of data-driven strategies in social sciences is not the end of theory; rather, it should be used to upgrade phenomenological pieces to a higher theoretical level that adapts to the social change.

References

1. Blei, D., & Lafferty, J. (2006). Correlated topic models. Advances in neural information processing systems, 18, 147.
2. Mattingly, D., & Yao, E. (2020). How Propaganda Manipulates Emotion to Fuel Nationalism: Experimental Evidence from China. Available at SSRN 3514716.
3. Mudde, C. (2004). The Populist Zeitgeist. Government and Opposition, 39(4), 541–563.

[3] Information receivers are users who receive populist information from others and respond, for example, by retweeting or mentioning it.

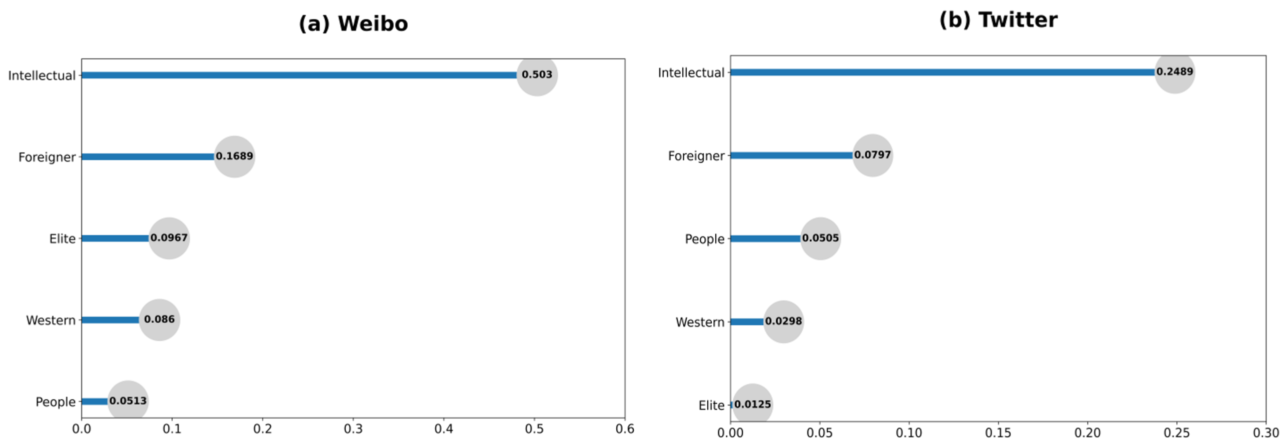

Figure 1: Proportion of Populist Posts by Social Media User Groups

Weibo

Category	Sub_topic
People	**Topic 1:** Popular sovereignty
	Topic 2: Corruption
	Topic 3: Human right
	Topic 4: Accountability
	Topic 5: Betrayal
Anti-foreigner	**Topic 1:** Betrayal
	Topic 2: Super national treatment
	Topic 3: Culture invasion
Anti-western	**Topic 1:** Immoral west
	Topic 2: Betrayal
Anti-intellectual	**Topic 1:** Betrayal
	Topic 2: Culture invasion
Anti-elite	**Topic 1:** Corruption
	Topic 2: Accountability

No.	Topic	No.	Topic
1	Popular sovereignty	5	Betrayal
2	Corruption	6	Super national treatment
3	Human right	7	Culture invasion
4	Accountability	8	Immoral west

Twitter

Category	Sub_topic
People	**Topic 1:** Democracy
	Topic 2: Corruption
	Topic 3: Party
Anti-foreigner	**Topic 1:** Super national treatment
	Topic 2: Betrayal
	Topic 3: Democracy
Anti-western	**Topic 1:** Immoral west
	Topic 2: Liberal democracy
Anti-intellectual	**Topic 1:** Betrayal
Anti-elite	**Topic 1:** Corruption
	Topic 2: Party

No.	Topic	No.	Topic
1	Democracy	5	Betrayal
2	Corruption	6	Immoral west
3	Party	7	Liberal democracy
4	Super national treatment		

Figure 2: Topics of Populist Posts on Weibo and Twitter

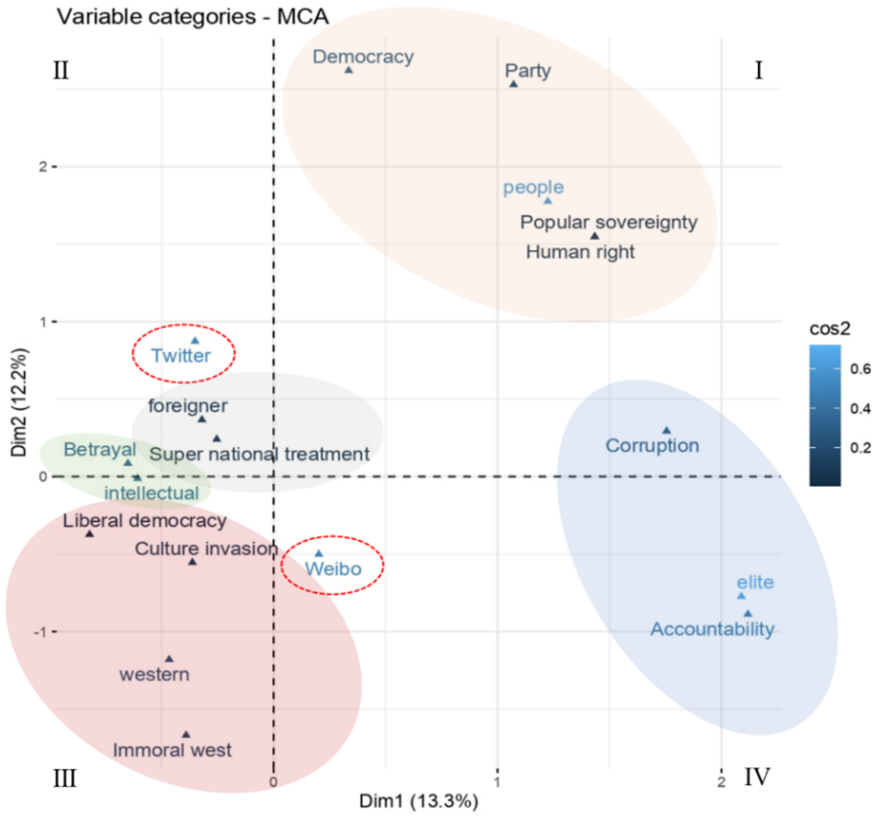

Figure 3: Semantic Dimensions Measured by Correlations (MCA was applied to the merged dataset of Weibo and Twitter. Dim 1 reflects the Destructive-Constructive scale; Dim 2 reflects the Populist-Ultranationalist scale.)

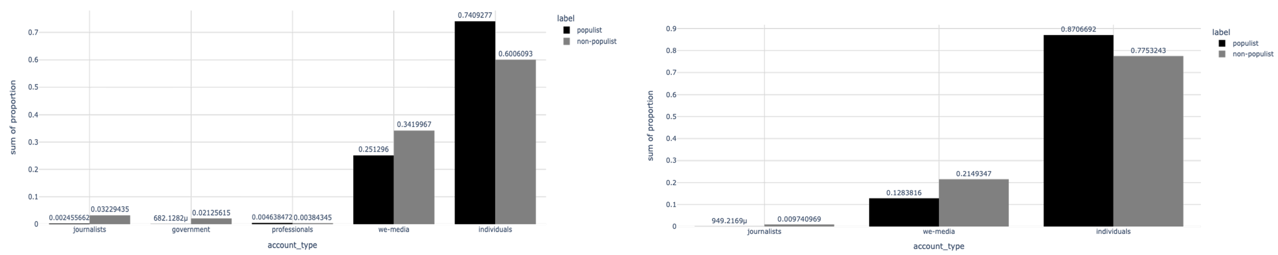

Figure 4: Semantic Dimensions Measured by Correlations (Distribution of Account Types on Weibo (left) and Twitter (right)

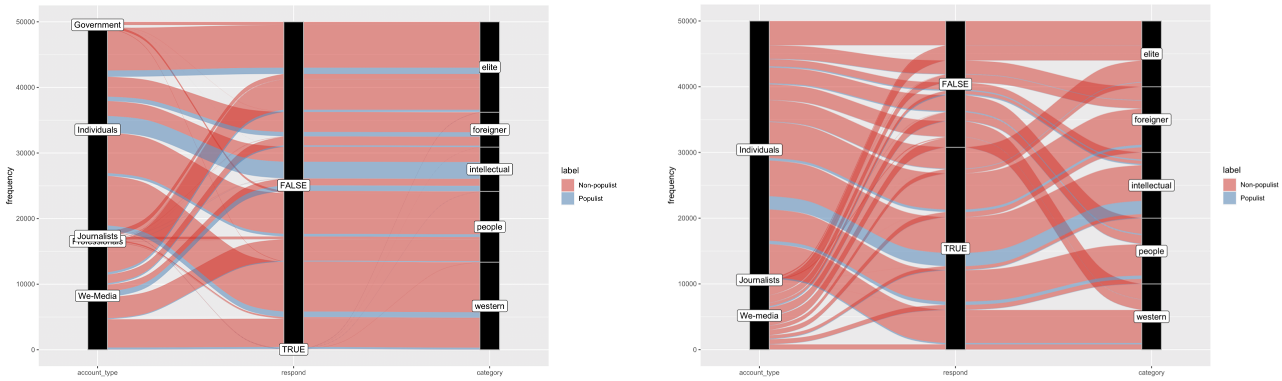

Figure 5: Parallel Plots of Account Types, Response Type, and Targets on Weibo (left) and Twitter(right)