Comparing Objectivity:
An Empirical Study on Group Bias in Clickbait Headlines across Different Media Outlets

Siyu Zhang
SiChuan International Studies University

CONTENTS

01
PART

Research Question and Hypothesis

Research Question and Hypothesis

News headlines and Clickbait

Clickbait is characterized as a type of headline primarily designed to capture users' attention and prompt them to click, often lacking adequate information (Chen et al., 2015). Based on the existing literature, further refinement is needed in the exploration of clickbait phenomena.

Firstly, most studies predominantly examine clickbait in legacy media contexts, while algorithm-driven recommendation platforms encompass a range of authors, including self-media and official organization media.

Secondly, current research primarily focuses on generalized news media, but within the Chinese context, there are varying layers even among news media types.

Research Question 1: Does the clickbait phenomenon vary across different author types?

Research Question and Hypothesis

Schema Resonance Model (SRM)

    This study utilizes the Schema Resonance Model (SRM) to understand the dynamic relationship between news organizations and the media. Schema refers to an individual's cognitive structure of knowledge about concepts or external stimuli (Arbib, 2003), which is closely related to personal experience (Taylor & Crocker, 1981; van Gorp, 2007). The SRM describes the resonance between the schemas utilized by the supply side in meaning production and the schemas employed by the receiver side in meaning interpretation. This model is often applied to understand the purchasing behaviors of merchants and customers in the economic domain (Song & Christen, 2019).

Research Question and Hypothesis

Schema Resonance Model (SRM)

We posit that clickbait headlines, designed to attract more clicks, tend to resonate with readers' mental frameworks by appealing to their pre-existing biases. As such, we propose the following research hypotheses and questions:

Hypothesis 1: Clickbait headlines exhibit greater negativity in group representations compared to traditional headlines.

Research Question 2: Do different authors produce headlines with varying biases in group representation?

02

Methodology

Data Collection

1. We simulates user scrolling behavior to collate all available article titles and their respective authors from the homepage

2. Collected article authors as the focal point for a secondary round of data gathering, capturing article titles, publication timestamps, read and like counts, in addition to the authors' number of fans and verification details from their user pages.
In total, the study amasses 103,773 article titles authored by 502 contributors.
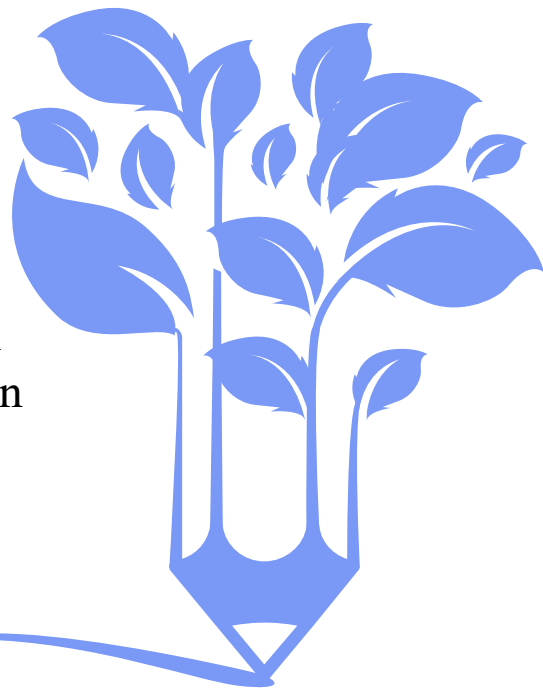
Data Collection

Clickbait Classifier

The study employed supervised machine learning techniques to categorize the collected article titles.
one thousand article titles were randomly selected from a pool of sample titles.

The study trained the headline classifier with feature variables based on the length of the headline, the number of commas, exclamation marks, and question marks et al.

Data Collection

Authors Type Categorization

The authors' identities were classified into five categories: unauthenticated self-media, authenticated self-media, other news media, authoritative news media, and institutional-owned media
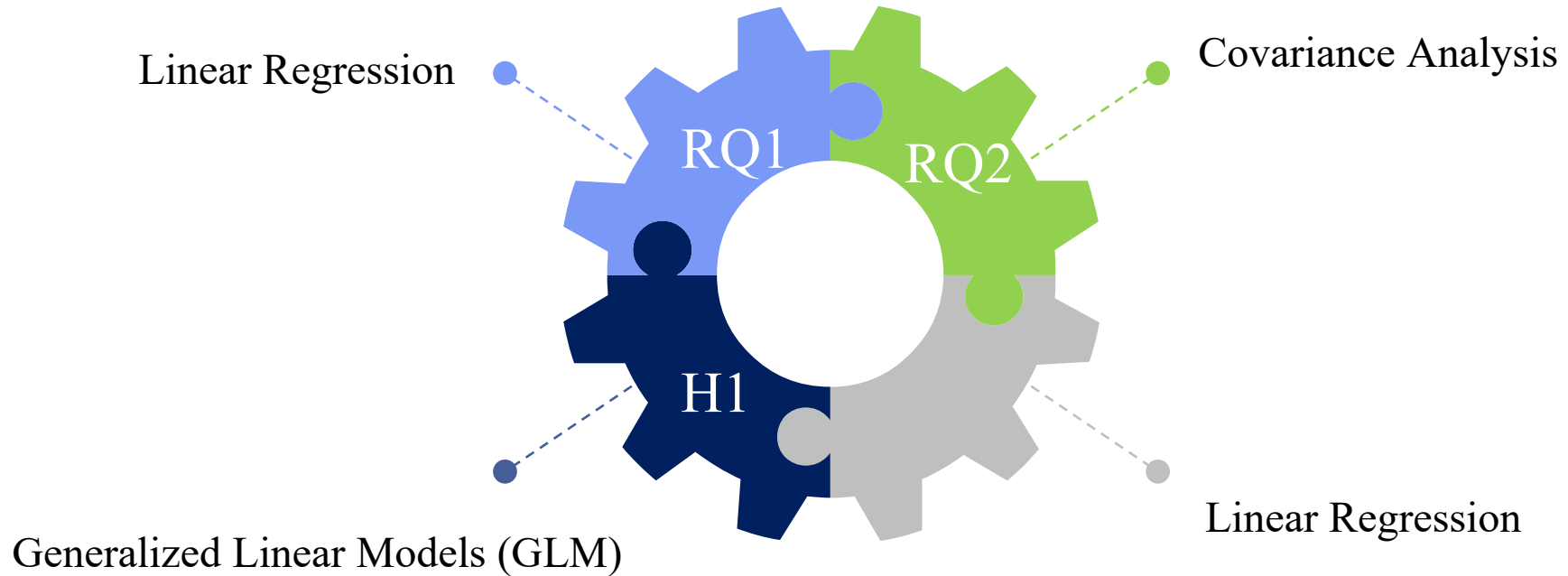
Data Collection

Network Analytics →

We selected article titles using gender (male, female), age (youth, middle-aged, elderly), household registration (urban, rural), and related synonyms. This resulted in a total of 7,667 titles containing group identity characteristics.

To measure bias in these titles, the researchers utilized the Perspective API. We conducted a principal component analysis (PCA) to construct a negative prejudice index.

Negative Bias Index = TOXICITY x 0.34 + INSULT x 0.3286 + PROFANITY x 0.3286

Analysis Strategy

Linear Regression

Covariance Analysis

RQ1

RQ2

H1

Generalized Linear Models (GLM)

Linear Regression

03
PART

Findings

Clickbait Probability

Unauthenticated self-media

37.28 %

Authenticated self-media

42.59 %

Other news media

30.31 %

Authoritative news media

13.28%

Institutional-owned media

38.15%

Negative Bias

Table 1. Covariance analysis of Author Types Negative Bias Index (N=7667)

Author Types	Numbers	Negative Bias Index Mean	SD
Unauthenticated self-media	3960	0.1213	0.1083
Authenticated self-media	1584	0.1187	0.1147
Other news media	1751	0.0819	0.0812
Authoritative news media	109	0.0336	0.0491
Institutional-owned media	263	0.0768	0.0926

Negative Bias

Table 2. Generalized linear Regression Analysis of Clickbait Probability and Negative Bias Index

Viable	Model1		Model2	
	Coeff	std err	Coeff	std err
Constant	0.1177***	0.002	0.1161***	0.003
Dummy virable1	-0.0031	0.003	0.0037	0.005
Dummy virable2	-0.0395***	0.003	-0.0369***	0.004
Dummy virable3	-0.0899***	0.011	-0.1009***	0.012
Dummy virable4	-0.0472***	0.007	-0.0474***	0.01
Publication data	-4.31E-06	7.06E-06	-4.31E-06	7.06E-06
The number of fans	3.13E-10	2.40E-10	2.50E-10	2.42E-10
Clickbait Probabality	0.0094***	0.003	0.0131***	0.004
Dummy virable1*Clickbait Probabality			-0.0146	0.008
Dummy virable2*Clickbait Probabality			-0.0062	0.008
Dummy virable3*Clickbait Probabality			0.0662*	0.029
Dummy virable4*Clickbait Probabality			0.0007	0.016
Log-Likelihood	6558.4		6563.4	
Pseudo R-squ. (CS)	0.03679		0.03807	
BIC	-68416.26988		-68380.59644	

The generalized linear regression analysis results, revealed a positive relationship between the probability of clickbait and the negative bias index of group representation.

Group Bias

1. Women mentioned nearly twice as many times as men in headlines. Headlines were also more likely to negatively characterize women in headline parties that mentioned them.

2. On the topic of age, youth was the most mentioned group, and middle-aged groups were more likely to be negatively characterized than youth and old age groups.

04

PART

Conclusion

Author type and Clickbait

3%

clickbait in traditional media headlines on Twitter

6%

clickbait in traditional media headlines on Facebook

13.28%

clickbait in traditional media headlines on Toutiao

This disparity is attributed to the market dynamics of domestic media, where the imbalance between information supply and demand, driven by the marketization of news and the rise of self-media, compels media organizations to seek external funding for survival.

Group Bias in headlines

Within the context of our research, women can be identified as a digitally disadvantaged group, utilized as tools to maximize media profits through the perpetuation of stereotypes.

We suggest that in the new media landscape driven by algorithms, traditionally disadvantaged groups continue to face or even exacerbate their disadvantaged position.

Group Bias in headlines

The discovery affirms the applicability of the schema resonance model within the field of information dissemination.

Media, as information providers, leverage individuals' existing schemas to create and distribute title parties effectively. This enables people to comprehend and engage with these titles based on their preconceived notions.

Comparing Objectivity:
An Empirical Study on Group Bias in Clickbait Headlines across Different Media Outlets

Siyu Zhang
SiChuan International Studies University