INFORMATICA, 2024, Vol. 35, No. 3, 671–686

DOI: https://doi.org/10.15388/24-INFOR554

A Temporal Variable-Scale Clustering Method on Feature Identification for Policy Public-Opinion Management

Ai WANG[1,*], Xuedong GAO[2], Mincong TANG[3]

[1] *University of Science and Technology Beijing, Beijing 100083, China*
[2] *School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China*
[3] *Industrial University of Ho Chi Minh City, Vietnam*
e-mail: b2113131@ustb.edu.cn, gaoxuedong@manage.ustb.edu.cn, mincongtang@iuh.edu.vn

Received: January 2024; accepted: April 2024

Abstract. The development of various digital social network platforms has caused public opinion to play an increasingly important role in the policy making process. However, due to the fact that public opinion hotspots usually change rapidly (such as the phenomenon of public opinion inversion), both the behaviour feature and demand feature of netizens included in the public opinion often vary over time. Therefore, this paper focuses on the feature identification problem of public opinion simultaneously considering the multiple observation time intervals and key time points, in order to support the management of policy-focused online public opinion. According to the variable-scale data analysis theory, the temporal scale space model is established to describe candidate temporal observation scales, which are organized following the time points of relevant policy promulgation (policy time points). After proposing the multi-scale temporal data model, a temporal variable-scale clustering method (T-VSC) is put forward. Compared to the traditional numerical variable-scale clustering method, the proposed T-VSC enables to combine the subjective attention of decision-makers and objective timeliness of public opinion data together during the scale transformation process. The case study collects 48552 raw public opinion data on the double-reduction education policy from Sina Weibo platform during Jan 2023 to Nov 2023. Experimental results indicate that the proposed T-VSC method could divide netizens that participate in the dissemination of policy-focused public opinion into clusters with low behavioural granularity deviation on the satisfied observation time scales, and identify the differentiated demand feature of each netizen cluster at policy time points, which could be applied to build the timely and efficient digital public dialogue mechanism.

Key words: public opinion, variable-scale clustering, education policy, temporal observation scale.

1. Introduction

As AI technology has created more attractive application scenarios in the social network domain (Leung, 2023; August *et al.*, 2020), not only the number of netizens but also their

*Corresponding author.

activity on multiple digital platforms has increased rapidly (Kuo *et al.*, 2021). That is followed by the emergence of a large amount of online public opinion events, widely disseminated across different platforms (Akhter *et al.*, 2021). On the one hand, online public opinion plays a significant role on the policy making process, since it contains lots of valuable netizens' demand and attitude information (Xiao and Wong-On-Wing, 2021; Li *et al.*, 2022b; Lili *et al.*, 2020). On the other hand, sometimes the policy itself will also cause heated discussion of online public opinions, that is defined as the policy(-focused) public opinion (Wang *et al.*, 2019).

Moreover, the phenomena of public opinion inversion has aroused wide attention in both academic and industrial fields (Yuan *et al.*, 2017; Tan and Hua, 2019). The public opinion inversion refers to netizens' attitude, emotions or points of view quickly reversed in the opposite direction over time (Zhang *et al.*, 2023; Yang, 2023), which is proved to be the result of the synergistic evolution between multiple subjects in the dissemination of public opinion (Zhao *et al.*, 2023; Ai *et al.*, 2023; Zhu, 2023). It can be seen that the time attribute is an important dimension in observing the state of public opinion, while different observation time intervals (temporal scales) might lead to obtain different netizens' features. In particular, for the policy-focused public opinion, some key observation time points, like the promulgation time of relevant policies, have a more significant impact on public opinion. Therefore, in order to manage online public opinion, it is crucial to timely and accurately identify netizens' features following dynamic observation time.

The variable-scale data analysis theory (VSDA) (Wang and Gao, 2022) is used to study the influence of different types of observation scales on decision-making results, by simulating the scale transformation process of decision makers. The framework of VSDA could be classified into main three stages. Firstly, select a single scale data set for scale transformation from the multi-scale data model. Then (taking the clustering data analysis task as an example) perform the cluster analysis on the current single scale data model, and evaluate the satisfaction of the clustering results. Finally, iteratively adjust the observation scale of single scale data model according to the evaluation results, until all the divided clusters meet the satisfaction standard of management scenarios (Wang and Gao, 2021a).

Since different data types (such as spatial data (Radosevic *et al.*, 2023), numerical data (Wang and Wang, 2021), categorical data (Lee and Jung, 2021), binary data (Shati *et al.*, 2023), etc.) are suitable for different data structures, the structure representation model of different observation scales in the VSDA (i.e. the scale space model) also has different connection modes between multiple scale hierarchies (Wang and Gao, 2022). Although Wang and Gao (2021b) propose a variable-scale dynamic clustering method that is capable of describing the timeliness characteristics of time-related observation scales (like the material inventory data of aerospace project), its numerical scale space model could only reflect the single perspective that the latest period data provides more significance during the decision-making process. It is unable to meet the jointly observation requirements of multiple time intervals and key time points on the management of policy-focused public opinion.

Therefore, this paper studies the feature identification problem of public opinion considering the multiple observation time intervals and policy time points simultaneously, based on the variable-scale data analysis theory. The main contributions of our research are summarized as follows:

- In order to characterize the multi-level analysis requirements of time dimensions for the policy-focused public opinion management, the temporal scale space model is established to describe all the candidate temporal observation scales, which are organized following the time points of relevant policy promulgation (policy time points).
- According to the proposed temporal scale space model above, the multi-scale temporal data model is established to represent netizens' different behaviours in the dissemination of public opinion under every temporal observation scale, instead of just keeping one maximum value of the original time series in the research work (Wang and Gao, 2021b).
- A temporal variable-scale clustering method (T-VSC) is put forward. Compared to the traditional variable-scale dynamic clustering method for numerical data, the proposed T-VSC enables to combine the subjective attention of decision-makers and objective timeliness of public opinion data together during the scale transformation process for netizens' demand feature identification.

The paper is organized as follows. Section 2 introduces the previous research works, including public opinion management and variable-scale clustering methods. Section 3 presents the main part methodology of our research in detail. A case study on the real public opinion dataset of the double-reduction education policy is described in Section 4. The paper is concluded in the last Section.

2. Literature Review

2.1. *Public Opinion Management*

Online public opinion has gained wide attention not only in the policy-making process of governmental departments (Li, 2021; Awad *et al.*, 2020), but also in optimizing strategic decisions of enterprises in various industries (Fortin and Cimon-Morin, 2023; Rogowski, 2023; Krause and Gahn, 2023). Due to the diversity of participants in public opinion discussion and the complexity of information dissemination channels, an efficient identification of netizens' features becomes one of the key factors to manage public opinion (Ma *et al.*, 2023). Previous research works mainly focus on two aspects: the identification of netizens' demand and behaviour features.

As for the demand feature identification, Yang (2024) puts forward a theme clustering algorithm to obtain the theme portraits of public opinion on sudden public events. He (2023) proposes a short text mining algorithm to analyse netizens' attitude and emotion through online public opinion. Meanwhile, for the behaviour feature identification, Lv *et al.* (2023) puts forth the speech behaviour classification algorithm for analysing the development tendency of public opinion. However, most of the research above observes netizens' features only at the most recent fixed time interval (fixed temporal scale).

When making the management decisions on the policy-focused public opinion, it is necessary to consider not only the objective timeliness of public opinion data, but also the subjective attention of decision makers at special time points, like the promulgation of relevant policies (Zhang and He, 2023). Hence, this paper studies the feature identification problem of public opinion under a variable-scale observation perspective on the time dimension.

2.2. *Variable-Scale Clustering Methods*

Decision hierarchy transformation is the most significant thinking feature of intelligent decision analysis, as well as one of the important ways to realize data granulation in the data space (Liu and Zhang, 2019; Leng *et al.*, 2018). The variable scale data analysis theory (VSDA) establishes an automatic data analysis hierarchy (observation scale) transformation mechanism by simulating the decision analysis hierarchy transformation process of managers, so as to achieve the balance between the quality and efficiency of data analysis in the decision process (Wang *et al.*, 2020).

As the time dimension is one of the most commonly utilized observation rulers for decision analysis process in both business and policy-making scenarios, previous studies build various scale space models with different characteristics to describe the time-related data observation scales for differentiated management demands.

For instance, considering the versatile materials inventory management context of aerospace project, Wang and Gao (2021b) propose the numerical scale space model (see Fig. 1) to represent the data feature that the latest period data earns more significance on the decision-making process, including the purchasing cycle, demand quantity, cost of sales observation ruler, etc. Fig. 1 depicts that with the increase of the time interval observation scales in the concept chain, the number of data observation scale value decreases gradually, which causes the value space to appear as a unimodal state of aggregation towards the nearest time point t_n. Table 1 further shows the maximum data value of different types of material objects under the multiple observation scales, based on the numerical

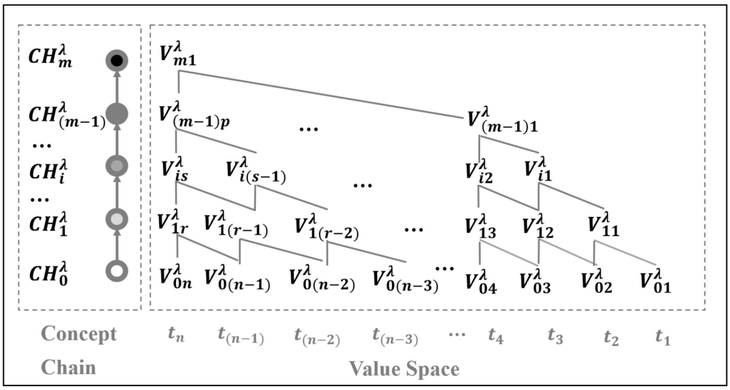

Fig. 1. Numerical scale space model of dimension (observation ruler) A^λ.

Table 1
Multi-scale numerical data model.

$\mathcal{U}$ \ A^S	A^{λ}				$A^{\lambda+1}$	
	A_0^{λ}	A_1^{λ}	A_3^{λ}	A_4^{λ}	$A_0^{\lambda+1}$	$A_1^{\lambda+1}$
x_1	V_{01}^{λ}	V_{11}^{λ}	$V_{(n-1)1}^{\lambda}$	V_{n1}^{λ}	$V_{01}^{\lambda+1}$	$V_{11}^{\lambda+1}$
x_2	V_{02}^{λ}	V_{11}^{λ}	$V_{(n-1)1}^{\lambda}$	V_{n1}^{λ}	$V_{01}^{\lambda+1}$	$V_{11}^{\lambda+1}$
x_3	V_{03}^{λ}	V_{11}^{λ}	$V_{(n-1)1}^{\lambda}$	V_{n1}^{λ}	$V_{02}^{\lambda+1}$	$V_{12}^{\lambda+1}$
x_4	V_{0j}^{λ}	V_{13}^{λ}	$V_{(n-1)1}^{\lambda}$	V_{n1}^{λ}	$V_{02}^{\lambda+1}$	$V_{12}^{\lambda+1}$
x_5	$V_{0(n-1)}^{\lambda}$	V_{1m}^{λ}	$V_{(n-1)2}^{\lambda}$	V_{n1}^{λ}	$V_{03}^{\lambda+1}$	$V_{12}^{\lambda+1}$
x_6	V_{0n}^{λ}	V_{1m}^{λ}	$V_{(n-1)2}^{\lambda}$	V_{n1}^{λ}	$V_{04}^{\lambda+1}$	$V_{13}^{\lambda+1}$

scale space model in Fig. 1, i.e. the numerical multi-scale data model. It can be seen that material objects could be divided into a smaller number of equivalence classes, following the scale up transformation of one observation ruler.

Furthermore, Wang *et al.* (2022) also study the space-time scale transformation problem on the charging and discharging behaviours of electric vehicle owners for the digital vehicle-to-grid (V2G) platform. Taking EV owners' different short-term demand response time intervals as temporal observation scales, as well as taking the distance intervals between EV owners to the target public charging pile at the V2G demand release time point as spatial observation scales, the space-time scale space model is established, where the value space shows the number of responses EV users provide under the fixed time and space observation scale. Although the space-time scale space model has paid attention to the importance of key time point and multiple time intervals when making scheduling strategies on EV owners, only one demand release time point is taken into consideration.

Therefore, this paper studies the netizens' feature identification problem for public opinion management based on the variable-scale data analysis, under both of the observation requirements on the data timeliness, as well as key time points.

3. Research Methods

Since the variable-scale data analysis theory (VSDA) has the advantage of modelling the multi-level decision analysis needs (Wang *et al.*, 2022), this section studies the feature identification problem of policy-focused public opinion while considering different types of data.

In order to describe the relation between multiple observation time intervals and key time points simultaneously, the temporal scale space model is established in Definition 1.

Definition 1 (*Temporal scale space model*). Given a time series $TS = (v_{t1}, v_{t2}, \dots, v_{tn})$ of observation ruler (dimension) A^{λ}, and v_t represents the value at time t, the temporal scale space model of A^{λ} is Temporal-$\boxed{\boxed{A^{\lambda}}} = \{CC, VS\}$, where the concept chain $CC = \{CH_k^{\lambda} | 0 \leqslant k \leqslant m\}$ and CH_k^{λ} is the kth observation scale of A^{λ}, the value space

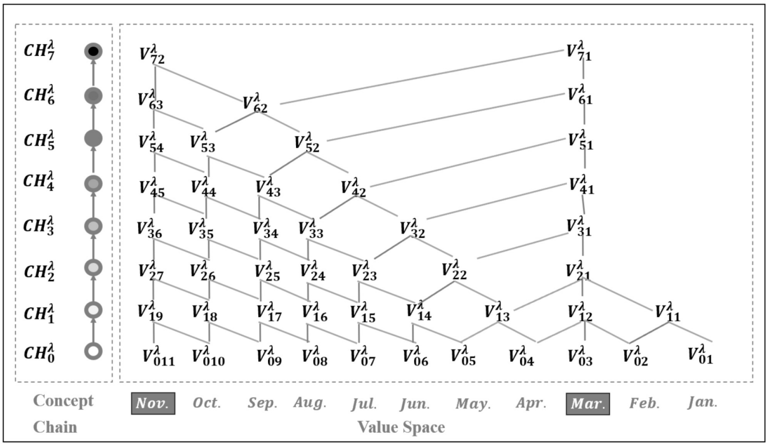

Fig. 2. Example: The temporal scale space model.

$VS = \{V^{\lambda}_{kt} | \, V^{\lambda}_{kt} = f(v_{(t-k):(t+k)}) \wedge (k+1 \leqslant t \leqslant n-k)\}$, if t is the policy time point (Observation time point); otherwise, $V^{\lambda}_{kt} = f(v_{(t-k):t}) \wedge (k+1 \leqslant t \leqslant n)$, f is the maximum information function, i.e. $f(v_{(t-k):(t+k)})$ means get the maximum value of TS in the time window $[t-k, t+k]$.

According to Definition 1, the temporal scale space model has the following properties: (1) The lower level temporal observation scale in the concept chain is partially ordered at the higher level scale; (2) The scale values in value space follow the partial order relationship between the scale hierarchies to which they belong.

Taking the public opinion on the double-reduction education policy as an example, the data collection process is shown in Section 4.1 in detail. Since there was an influential double-reduction relevant policy promulgated in March 2023, the above policy time point gains more attention from policymakers and the public. Moreover, combining the data timeliness requirement (that the latest period data earns more significance on the decision-making process, Wang and Gao, 2021b), March and November are the two key observation time points for netizens' feature identification. Hence, according to the Definition 1, the temporal scale space model could be built (see Fig. 2). It can be seen that with the increase of the time observation scale, the number of observation scale value in the value space becomes smaller and gathers two key points towards Mar. and Nov., showing a double-peak pattern.

According to the construction procedures of traditional scale space model (Wang and Gao, 2022), the temporal scale space model could be built mainly through three stages below.

On the first stage, determine all the candidate temporal scale hierarchies (time intervals) and clarify the key time points for the specific management scenario. For example, the basic (lowest) observation scale CH^{λ}_{0} in Fig. 2 is equal to the initial monthly interval in the double-reduction education policy case, while the adjacent higher scale CH^{λ}_{1} means the last two months.

On the second stage, correlate scale values according to the scale hierarchies from low to high, and follow the order of key time points within the same scale level, which could be broken down into three steps. (1) Extract the maximum value within the observation time interval at the latest time point, that is November in Fig. 2. (2) Extract the maximum value within the observation time interval at other key time points in a chronological order from far to near, that is March in Fig. 2. (3) Extract the maximum value within the observation time interval of the remaining time points in a chronological order from near to far.

At the last stage, reduce the current temporal scale space model from the high scale hierarchy to the low level, until its peak number is the same as the number of key time points.

Compared to the traditional numerical scale space model in the research of Wang and Gao (2021b), the proposed temporal scale space model keeps the scale hierarchies of the latest observation time intervals, while also emphasizes the influence of policy time points on netizens' participation behaviour in the public opinion dissemination, which provides a kind of problem-solving space (Allen and Simon, 1972) for the subsequent variable scale data analysis process.

After establishing the temporal scale space model, the multi-scale temporal data model is proposed in Definition 2, in order to comprehensively present various netizens' behaviour in the dissemination of public opinion under every temporal observation scale.

Definition 2 (*Multi-scale temporal data model*). Let Temporal-$D^S = (\mathcal{U}, \mathcal{A}^S, \mathcal{V}^S, f)$ represent the multi-scale temporal data model, where $\mathcal{U} = \{x_1, x_2, \dots, x_p\}$ is the object set (universe), $\mathcal{A}^S = \{A^1, A^2, \dots, A^r\}$ represents the observation attribute (observation ruler) set, where at least one attribute within $\mathcal{A}^S$ has multiple temporal scales in its temporal scale space, i.e. $\exists A^\lambda$, $CC(A^\lambda) = \langle CH_0^\lambda, CH_1^\lambda, \dots, CH_m^\lambda \rangle (A^\lambda \in \mathcal{A}^S)$, $f : \mathcal{U} \times \mathcal{A}^S \to \mathcal{V}^S$ is the information function, and $\mathcal{V}^S \in VS(A^\lambda)$, $A^\lambda \in \mathcal{A}^S$.

Hence, compared to the traditional multi-scale numerical data model in Table 1, the multi-scale temporal data model could represent netizens' different behaviour under every temporal observation scale (like the various behaviour features V_{0i}^λ $(i = 1, 2, \dots, 11)$ on the basic observation scale CH_0^λ in Fig. 2), instead of just keeping one maximum value of the original time series.

According to the temporal scale space model (Temporal-$\boxed{\boxed{A^\lambda}}$) and multi-scale temporal data model (Temporal-D^S), the mechanism of the temporal scale transformation (Temporal-ST) is proposed in Fig. 3.

In order to achieve the public opinion management for different types of netizens, the temporal scale transformation mechanism starts with the initial clustering process on the basic temporal scale of Temporal-D^S, which aims to obtain netizen clusters with similar behaviour feature. The measurement granular deviation GrD (see Eq. (1)) is utilized to evaluate whether the scale feature of each cluster satisfies decision requirements. According to the variable-scale data analysis theory (see Section 2.2), the satisfaction judgement standard of granular deviation is specified as the maximum granular deviation value of initial qualified clusters which are determined by decision makers. Hence, all the clusters

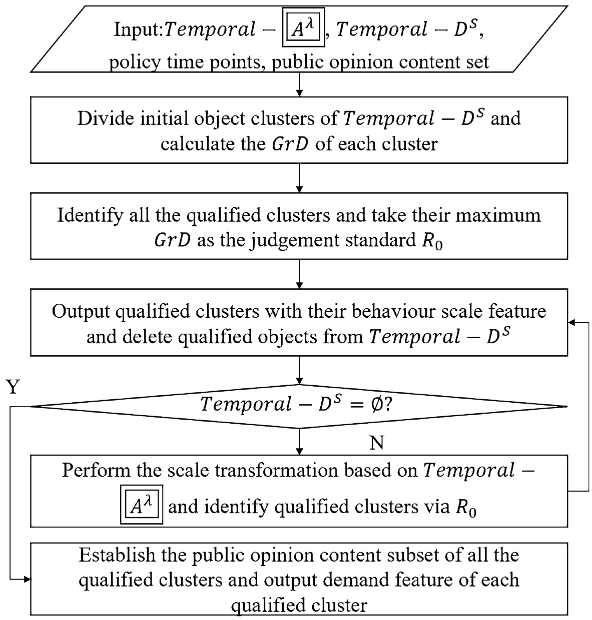

Fig. 3. The mechanism of the temporal scale transformation (*Temporal-ST*).

with larger granular deviation are unqualified clusters and need further scale transformation.

For the scale transformation process, such as improving the scale hierarchy from the lower level to higher level, let different observation values of qualified objects' equivalence classes on the target level be the equivalent interval (Wang and Gao, 2022) and replace it with the intermediate value. And then iteratively perform clustering for the remaining objects based on the Temporal-$\boxed{A^\lambda}$.

According to the temporal scale transformation mechanism *Temporal-ST*, a temporal variable-scale clustering method (T-VSC) is put forward, and the calculation steps are shown in Algorithm 1.

The time complexity of the method T-VSC is $O(t(\varphi + p))$, where t is the time complexity of the meta clustering method, $\varphi = \min(p, m^r)$, p is the number of netizens, r is the number of observation rulers of netizens' behaviour and m is the maximum number of temporal scale hierarchies in one ruler.

4. Results and Discussion

4.1. *Experiment Design and Data Collection*

The double-reduction policy, that aims to reduce students' homework burden and off-campus training burden, has attracted wide attention since it was first put forward in July 2021 (Li *et al.*, 2022a). Several new social phenomena caused by the implementation process of the double-reduction policy has become hot topics of online public opinion.

Algorithm 1 *Temporal variable-scale clustering* (policy time points, Temporal-$\boxed{\boxed{A^\lambda}}$, Temporal-$D^S$, initial evaluation dimension A^λ, public opinion content set) // Temporal-$D^S = (\mathcal{U}, \mathcal{A}^S, \mathcal{V}^S, f)$ is the multi-scale temporal data model of all the netizens; Temporal-$\boxed{\boxed{A^\lambda}}$ $(A^\lambda \in \mathcal{A}^S)$ are the temporal scale space model of netizens' behaviour observation rulers (while the data collection process is shown in Section 4.1 in detail).

1: Divide all objects in the universe $\mathcal{U}$ into initial clusters $X_I (I \in N^+)$ using the meta clustering algorithm (Wang and Gao, 2022) on the basic observation scale of Temporal-D^S.

2: Calculate the granular deviation GrD of all the initial clusters on A^λ.

$$GrD(X_I, A^\lambda) = \frac{\sum_{j=1}^{rm} (q_I - |x_{Ij}|)}{\frac{\sum_{k=1}^{s} |U_k^\lambda|^2}{|p|^2}}. \tag{1}$$

Where x_I is the centre of cluster X_I, q_I is the number of objects in X_I, $\mathcal{U}/A^\lambda = U_1^\lambda, U_2^\lambda, \dots, U_s^\lambda$, p is the number of objects in $\mathcal{U}$, r is the number of attributes in $\mathcal{A}^S$, and m is the number of scale hierarchy of Temporal-$\boxed{\boxed{A^\lambda}}$ $(A^\lambda \in \mathcal{A}^S)$.

3: Identify all the qualified initial clusters and take their maximum GrD as the judgement standard R_0.

4: Output qualified clusters with their behaviour scale feature, and delete all the objects in qualified clusters from Temporal-D^S.

5: If there are still objects left in Temporal-D^S (that is Temporal-$D^S \neq \phi$), go to step 6; otherwise, go to Step 8.

6: Perform the scale transformation of A^λ via Temporal-$\boxed{\boxed{A^\lambda}}$, and divide the rest objects into clusters using the meta clustering algorithm (Wang and Gao, 2022).

7: Take all the newly divided clusters whose GrD are lower than R_0 as qualified clusters, and go to Step 4.

8: Determine the observation time intervals via the policy time points in Temporal-$\boxed{\boxed{A^\lambda}}$, as well as the current observation scale of Temporal-D^S, and establish the public opinion content subset of each satisfied cluster.

9: Output the demand feature of each public opinion content subset for qualified cluster using the hybrid variable-scale clustering algorithm (Wang and Gao, 2019).

Timely and accurately identifying the core demands of different types of netizens through online public opinion could support policymaking in the next stage.

Numerical experiments in this section aim to verify the effectiveness of the proposed temporal variable-scale clustering method (T-VSC) on the feature identification scenarios for the policy public opinion management.

Since the year 2023 is a crucial phase for the implementation of the double-reduction education policy, we collect relevant public opinion data from January 2023 to November

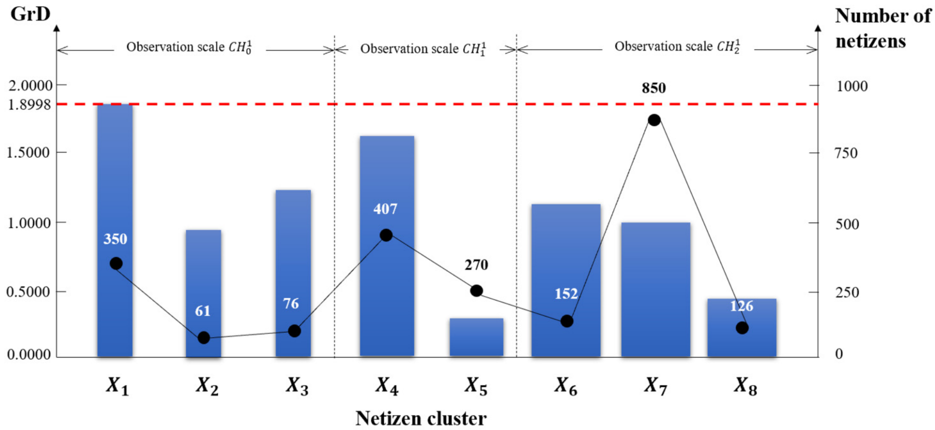

Fig. 4. The temporal scale transformation process on the behaviour feature identification.

2023 from the Sina Weibo social network platform on the monthly basic observation scale, and obtain a total of 48552 raw data.

In order to reflect the behaviour feature of netizens, the frequency of original content publishing (A^1), frequency of forwarding and commenting (A^2), as well as frequency of being liked (A^3) are taken as observation rulers. For the temporal scale space model of each observation ruler, the temporal observation scale CH_0^1, CH_1^1 and CH_2^1 are, respectively, the latest one month, the latest two months, the latest three months. March is the first policy time point t_I due to the promulgation of the double-reduction related policy. Moreover, considering the data timeliness requirement, the last month November is regarded as the second observation time point t_{II}.

4.2. *Experiment Results and Discussion*

According to the temporal variable-scale clustering method (T-VSC) in Algorithm 1, we identify netizens' feature from the behaviour and demand – two aspects, through the collected raw public opinion data on the double-reduction education policy (see Section 4.1). During the data preprocessing, we track the raw data of 2292 netizen users, who have passed the identity authentication by the platform, as well as participated in the dissemination of public opinion throughout January until November 2023.

Fig. 4 shows the temporal scale transformation process by the proposed method T-VSC. In Fig. 4, the height of rectangles represents the granular deviation value (*GrD*) of every netizen cluster, while the width of rectangles represents the scale hierarchy of temporal observation rulers and higher observation scales have larger width; the dotted line means the satisfaction judgement standard of *GrD*, and the broken line is the number of netizens in every cluster. It can be seen that all netizens are divided into eight clusters with satisfied scale feature through performing the scale up transformation twice.

Comparative experiments are conducted between the traditional single scale clustering method (SSC) (Wang and Gao, 2022) and the proposed T-VSC, and the evaluation results

Table 2
Comparative experimental results between the proposed T-VSC and the traditional single scale clustering method SSC.

Evaluation results of granular deviation index		SSC			T-VSC
		Observation scale CH_0^1	Observation scale CH_1^1	Observation scale CH_2^1	
	X_1	1.5149	0.0000	0.0000	1.8998
	X_2	1.5802	1.0943	1.0729	0.9737
	X_3	1.7756	1.6132	1.1415	1.2471
Netizen Cluster	X_4	1.7810	1.7899	1.2923	1.5701
	X_5	1.9934	1.8002	1.4431	0.2556
	X_6	2.0012	1.8898	1.4709	1.1842
	X_7	2.2319	2.1090	1.9806	1.0492
	X_8	2.2857	2.1864	2.0617	0.4718
Average evaluation result		1.8955	1.5604	1.3079	1.0814
Maximum evaluation result		2.2857	2.1864	2.0617	1.8998

Table 3
Experiment results of netizens' behaviour feature obtained by the proposed method T-VSC.

Netizen cluster			X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Scale hierarchy			Observation scale CH_0^1			Observation scale CH_1^1		Observation scale CH_2^1		
Netizens' behaviour feature under variable time intervals	A^1: Frequency of original content publishing	max	75	**449**	141	25	16	14	18	40
		min	0	0	0	3	1	1	0	0
		avg	25.91	112.18	54.73	10.36	3.82	3.82	2.27	8.45
		std	20.96	130.01	56.43	**5.73**	4.45	3.64	5.05	11.3
	A^2: Frequency of forwarding and commenting	max	0	19	**34**	5	1	12	18	4
		min	0	1	0	0	0	1	0	0
		avg	**0**	3.09	3.82	0.45	0.09	2.73	2.18	0.73
		std	0	5.07	9.59	**1.44**	0.29	3.02	5.08	1.21
	A^3: Frequency of being liked	max	43	120946	6255	2	127207	80445	1755	**200268**
		min	23	110946	76	0	125909	127	1755	199743
		avg	36.33	117606	2142	0.67	126774.33	53506	1755	200093
		std	9.43	4709.34	2908	**0.94**	611.88	**37745**	**0**	247.49

are shown in Table 2 in detail. It can be seen that although the granular deviation evaluation results of the SSC become smaller as the scale hierarchy increases, the average and maximum evaluation results of the SSC are still larger than the proposed T-VSC, which further verifies the efficiency of the T-VSC method.

Moreover, Tables 3 and 4 further describes the behaviour and demand feature of every netizen cluster under certain temporal observation scales, as well as observation time points.

There are three qualified netizen clusters (that is X_1, X_2, X_3) obtained by the T-VSC on the basic observation scale the latest one month. Among them, cluster X_2 earns the most frequency of original content publishing, reaching 499 in one month, but its standard deviation is also relatively large as 130.01. That means the publishing behaviour of X_2 stays in a highly fluctuating state. Referring to the demand feature in Table 4, netizens in X_2 mostly care about the international school education and head training institutions

Table 4
Experiment results of netizens' demand feature obtained by the proposed method T-VSC.

Netizen cluster	GrD	Number of netizens	Scale hierarchy	Netizens' demand feature at policy time points	
				Observation time point t_I	Observation time point t_{II}
X_1	1.8998	350	Observation scale CH_0^1	Qualification of training institutions	After-school service design; Teachers' competence and teaching skills
X_2	0.9737	61		International school education	New regulations of Beijing high school entrance examination; Ways of physical training
X_3	1.2471	76		Head training institutions Qualification of training institutions; Education bureau	Transformation of learning machine market
X_4	1.5701	407	Observation scale CH_1^1	Homework design; Teaching effect of in-school courses	Teachers' workload; Students' examination scores
X_5	**0.2556**	270		Pre-charge phenomenon of training courses	Regulation of teachers' flexible working schedule
X_6	1.1842	152	Observation scale CH_2^1	Science and technology training programs; Homework design	Students' career planning; Students' learning interest
X_7	1.0492	850		Art training programs; Teachers' competence and teaching skills	After-school service design; Off-campus training course management
X_8	0.4718	126		Students' sleep and eyesight protection; English training	Youth employment problem; Students' self-confidence

at observation time point t_I. Gradually, their attention has shifted to new regulations of Beijing high school entrance examination and ways of physical training over time. The above features indicate that although number of netizens in X_2 is small, they are relatively sensitive to education policies. Netizen cluster X_3 owns the maximum frequency of forwarding and commenting, while only cluster X_1 has never forwarded or commented any content during the whole time span. But both of them care about the qualification of training institutions at t_I shown in Table 4.

Taking the satisfaction judgement standard R_0 as the maximum GrD of the initial three clusters that is 1.8998 (shown in Fig. 4), the first scale up transformation process obtains two qualified netizen clusters X_4 and X_5. The standard deviation of three behaviour dimensions of cluster X_4 is all at a low level, which implies the behaviour feature of X_4 is relatively stable. And the demands of the second largest cluster X_4 are also quite consistent at two time points, including teaching effect of in-school courses, students' examination scores, homework design, etc. Cluster X_5 with the smallest behavioural granular deviation pays close attention to the pre-charge phenomenon of training courses at t_I and regulation of teachers' flexible working schedule at t_{II}.

Finally, three netizen clusters (i.e. X_6, X_7, X_8) are obtained on the latest three months observation scale. Cluster X_6 and X_7, respectively, get the largest and smallest standard deviation on the frequency of being liked, while cluster X_8 gets the most likes, exceeding two hundred thousand. According to Table 4, most of them care about the science and technology training programs, art training programs, students' career planning and employment related issues. These netizens' features could be applied to build the timely and efficient digital public dialogue mechanism.

5. Conclusions

In this paper, we address the feature identification problem for the management of policy-focused public opinion. According to the variable-scale data analysis theory, the research starts from establishing the temporal scale space model considering the influence of the data timeliness and key observation time points on netizens' feature identification process. The multi-scale temporal data model is established with the aim to represent netizens' different behaviours in the dissemination of public opinion on the basis of the temporal scale space model.

After proposing the temporal scale transformation mechanism, a temporal variable-scale clustering method (T-VSC) is put forward. Compared to the traditional numerical variable-scale clustering method, the proposed T-VSC enables to combine the subjective attention of decision-makers and objective timeliness of public opinion data together during the scale transformation process. The efficiency of the proposed method T-VSC is verified by 48552 real public opinion data on the double-reduction education policy from the Sina Weibo platform. Experimental results indicate that the proposed T-VSC method could divide netizens that participating in the dissemination of policy-focused public opinion into clusters with low behavioural granularity deviation on the satisfied observation time scales, and identify the differentiated demand feature of each netizen cluster at policy time points.

In the future, we will keep studying the real-time decision-making constraint in intelligent management scenarios on the mechanism of temporal scale transformation process.

Acknowledgements

The research was supported by China Postdoctoral Science Foundation (2021M700390), National Natural Science Foundation of China (71272161), China Scholarship Council (201906460087), and the Fundamental Research Funds for the Central Universities, China Grand number: FRF-TP-22-132A1.

References

Ai, J., Geng, L., An, Y., Hu, Z. (2023). Research on the prediction of network public opinion reversal based on relevance vector machine. *Computer Era*, (05), 113–117. https://doi.org/10.16644/j.cnki.cn33-1094/tp.2023.05.025.

Akhter, M., Jiangbin, Z., Naqvi, S.I., Abdelmajeed, M., Zia, T. (2021). Abusive language detection from social media comments using conventional machine learning and deep learning approaches. *Multimedia Systems*, 28(6), 1925–1940. https://doi.org/10.1007/s00530-021-00784-8.

Allen, N., Simon, H.A. (1972). *Human Problem Solving*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

August, T., Pescott, O., Joly, A., Bonnet, P. (2020). AI naturalists might hold the key to unlocking biodiversity data in social media imagery. *Patterns*, 1(7), 100116. https://doi.org/10.1016/j.patter.2020.100116.

Awad, E., Anderson, M., Anderson, S., Liao, B. (2020). An approach for combining ethical principles with public opinion to guide public policy. *Artificial Intelligence*, 287, 103349. https://doi.org/10.1016/j.artint.2020.103349.

Fortin, D., Cimon-Morin, J. (2023). Public opinion on the conflict between the conservation of at-risk species and the extraction of natural resources: the case of caribou in boreal forest. *The Science of the Total Environment*, 897, 165433. https://doi.org/10.1016/j.scitotenv.2023.165433.

He, H. (2023). Text mining and emotion analysis of online public opinion based on short review data of film websites. *Modern Information Technology*, 7(21), 126–135. https://doi.org/10.19850/j.cnki.2096-4706.2023.21.029.

Krause, W., Gahn, C. (2023). Should we include margins of error in public opinion polls? *European Journal of Political Research*. https://doi.org/10.1111/1475-6765.12633.

Kuo, Y.-F., Hou, J.-R., Hsieh, Y.-H. (2021). The advertising communication effectiveness of using netizen language code-switching in Facebook ads. *Internet Research*, 31(5), 1940–1962. https://doi.org/10.1108/INTR-04-2020-0231.

Lee, C., Jung, U. (2021). Context-based geodesic dissimilarity measure for clustering categorical data. *Applied Sciences*, 11(18), 8416. https://doi.org/10.3390/app11188416.

Leng, J., Chen, Q., Mao, N., Jiang, P. (2018). Combining granular computing technique with deep learning for service planning under social manufacturing contexts. *Knowledge-Based Systems*, 143, 295–306. https://doi.org/10.1016/j.knosys.2017.07.023.

Leung, R. (2023). Using AI-ML to augment the capabilities of social media for telehealth and remote patient monitoring. *Healthcare*, 11(12), 1704. https://doi.org/10.3390/healthcare11121704.

Li, J., Yan, S., Zhang, X., Li, X. (2022a). The media public opinion analysis on the implementation of "Double Reduction" policy in education based on big data. *Wireless Communications and Mobile Computing*, 2022, 1–10. https://doi.org/10.1155/2022/1093358.

Li, X., Li, Q., Du, Y., Yongquan, F., Chen, X., Shen, F., Xu, Y. (2022b). A novel tripartite evolutionary game model for misinformation propagation in social networks. *Security and Communication Networks*, 2022, 1–13. https://doi.org/10.1155/2022/1136144.

Li, Z. (2021). Forecast and simulation of the public opinion on the public policy based on the Markov model. *Complexity*, 2021, 1–11. https://doi.org/10.1155/2021/9936965.

Lili, D., Lei, S., Gang, X. (2020). Public opinion analysis of complex network information of local similarity clustering based on intelligent fuzzy system. *Journal of Intelligent & Fuzzy Systems*, 39(2), 1–8. https://doi.org/10.3233/JIFS-179943.

Liu, H., Zhang, L. (2019). Advancing ensemble learning performance through data transformation and classifiers fusion in granular computing context. *Expert Systems with Applications*, 131, 20–29. https://doi.org/10.1016/j.eswa.2019.04.051.

Lv, X., Dong, L., Teng, S., Zhang, L. (2023). Dataset construction and recognition method of multimodal language public opinion. *Journal of Beijing Information Science & Technology University*, 38(5), 1–9. https://doi.org/10.16508/j.cnki.11-5866/n.2023.05.001.

Ma, N., Yu, G., Jin, X., Zhu, X. (2023). Quantified multidimensional public sentiment characteristics on social media for public opinion management: evidence from the COVID-19 pandemic. *Frontiers in Public Health*, 11, 1097796. https://doi.org/10.3389/fpubh.2023.1097796.

Radosevic, N., Duckham, M., Rahaman, M., Ho, S., Williams, K., Hashem, T., Tao, Y. (2023). Spatial data trusts: an emerging governance framework for sharing spatial data. *International Journal of Digital Earth*, 16(1), 1607–1639. https://doi.org/10.1080/17538947.2023.2200042.

Rogowski, J. (2023). Public opinion and presidents' unilateral policy agendas. *American Journal of Political Science*, 67(4), 1134–1150. https://doi.org/10.1111/ajps.12753.

Shati, P., Cohen, E., McIlraith, S. (2023). SAT-based optimal classification trees for non-binary data. *Constraints*, 28(2), 166–202. https://doi.org/10.1007/s10601-023-09348-1.

Tan, Y., Hua, C. (2019). Analysis of cluster on the inversion problem of network public opinion events. *Journal of Yunnan University (Natural Sciences Edition)*, 41(S1), 16–20. https://doi.org/CNKI:SUN:YNDZ.0.2019-S1-003.

Wang, A., Gao, X. (2019). Hybrid variable-scale clustering method for social media marketing on user generated instant music video. *Tehnicki Vjesnik*, 26(3), 771–777. https://doi.org/10.17559/TV-20190314152108.

Wang, A., Gao, X. (2021a). A variable scale case-based reasoning method for evidence location in digital forensics. *Future Generation Computer Systems*, 122, 209–219. https://doi.org/10.1016/j.future.2021.03.019.

Wang, A., Gao, X. (2021b). A variable-scale dynamic clustering method. *Computer Communications*, 171, 163–172. https://doi.org/10.1016/j.comcom.2021.03.009.

Wang, A., Gao, X. (2022). *Variable-Scale Data Analysis Theory*. Economic Science Press, Beijing.

Wang, A., Gao, X., Tang, M. (2020). Computer supported data-driven decisions for service personalization: a variable-scale clustering method. *Studies in Informatics and Control*, 29(1), 55–65. https://doi.org/10.24846/v29i1y202006.

Wang, A., Gao, X., Tang, M. (2022). A space variable-scale scheduling method for digital vehicle-to-grid platform under distributed electric energy storage. *Applied Soft Computing*, 133, 109911. https://doi.org/10.1016/j.asoc.2022.109911.

Wang, H., Wang, H. (2021). Differentially private publication for correlated non-numerical data. *The Computer Journal*, 65(7), 1726–1739. https://doi.org/10.1093/comjnl/bxab014.

Wang, Y., Ning, J., Xubu, M. (2019). Forecast of public opinion on public policy based on social media. *Information Studies: Theory & Application*, 42(01), 87–93. https://doi.org/CNKI:SUN:QBLL.0.2019-01-015.

Xiao, F., Wong-On-Wing, B. (2021). Employee sensitivity to the risk of whistleblowing via social media: the role of social media strategy and policy. *Journal of Business Ethics*, 181(2), 519–542. https://doi.org/10.1007/s10551-021-04914-0.

Yang, C. (2023). Impact of multi-source perception of emergency events on reversal of public opinion from perspective of block modal coupling. *Journal of Intelligence*, 1–8. http://kns.cnki.net/kcms/detail/61.1167.G3.20231225.0950.012.html.

Yang, Y. (2024). Theme portrait and attribution analysis of public appeal for sudden public events from a data-driven. *Information Studies: Theory & Application*, 1–10.

Yuan, Y., Lan, Y., Zhang, P., Xia, Y. (2017). Research on the classification and forecast of reversal network public opinion based on cluster analysis. *Information Science*, 35(9), 54–60. https://doi.org/10.13833/j.cnki.is.2017.09.009.

Zhang, G., He, H. (2023). The evolutionary logic of public communication on policy in digital media: systematic investigation based on the temporal and spatial dimension. *Global Journal of Media Studies*, 10(3), 138–152.

Zhang, J., Zhang, P., Lan, Y., Zhong, Y. (2023). Construction and empirical study of online public opinion inversion identification model based on dynamic topic clustering. *Information Studies: Theory & Application*, 46(10), 174–181. https://doi.org/10.16353/j.cnki.1000-7490.2023.10.022.

Zhao, L., Wen, G., Yang, Y. (2023). PCA-LDA-LSSVM model for predicting network public opinion reversal of emergencies. *Journal of Safety Science and Technology*, 19(8), 186–190.

Zhu, G. (2023). Derived risks and their transmission in internet public opinion reversal: type division and relief strategy. *Journal of Jishou University (Social Sciences)*, 44(03), 99–111. https://doi.org/10.13438/j.cnki.jdxb.2023.03.009.

A. Wang received her PhD degree in 2021 in management science and engineering and she is currently an associate professor in the School of Humanities and Social Science at University of Science and Technology Beijing, China. She has published papers in respected journals like *Future Generation Computer Systems*, *Applied Soft Computing*, *Studies in Informatics and Control*. Her research interests focus on data mining, intelligent decision making, as well as emergency management.

X. Gao received his bachelor degree from Nankai University, China, in 1983, and the PhD degree from Belarusian State University in 1993. He is currently a professor in the Department of Management Science and Engineering, School of Economics and Management at University of Science and Technology Beijing, China. His research interests include management process optimization, data mining, intelligent decision making.

M. Tang graduated from The Chinese University of Hong Kong in 2011 with a PhD in management information systems. His main research areas include information management and systems, artificial intelligence, modelling and simulation in operations and transportation. Dr. Tang currently serves as an editor for several journals, including the *International Journal of Computers Communications and Control*, *International Journal of RF Technologies*, *Journal of Computing and Information Technology*. He has previously worked as a researcher at the International Research Center for Informatics Research at Beijing Jiaotong University and is now a chair professor at Xuzhou University of Technology, visiting professor at the Industrial University of Ho Chi Minh City in Vietnam. His research has been published in numerous international journals, including *IEEE Transactions on Fuzzy Systems*, *IEEE Transactions on Consumer Electronics*, *Eletronic Markets*, *Advances in Production Engineering and Management*, *Information Technology and Management*, *Information and Management*, *Journal of Applied Research and Technology*, *International Journal of Computers*, *Communications and Control*, *Studies in Informatics and Control*, *Applied Soft Computing*, *Future Generation Computer Systems*.