Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-05638-6 OPEN

A human-machine collaborative dynamic group consensus mechanism for mitigating manipulative tendencies

Yuzhou Hou[1], Xuanhua Xu[1,2]✉, Zongrun Wang[1] & Weiwei Zhang[2]

With the increasing frequency of emergencies, reliable public opinion fusion has become an important research topic in public opinion analysis and management. However, the public is unorganized and susceptible to manipulation, which poses a challenge. Therefore, from the perspective of coevolution, a human-machine collaborative decision-making mechanism considering manipulative behavior is constructed to ensure the timeliness, democracy, and reliability of public opinion fusion in emergencies. First, an opinion-trust coevolution process is proposed to simulate the human group decision-making environment. Next, a function of the degree of manipulation tendency is constructed based on the extreme opinion expression and influence behaviors of individuals. Then, a machine moderator is trained to manage manipulative behaviors via the feedback adjustment parameters of the human group's social network, and a human-machine collaborative decision-making mechanism is constructed. Finally, the proposed method is applied to public opinion fusion by considering a torrential rainstorm in Fujian Province, China, as a case study. The results of simulation analyses verify the reliability and effectiveness of the proposed mechanism.

[1] Central South University, Changsha, China. [2] Hunan University of Technology and Business, Changsha, China. ✉email: xuxh@csu.edu.cn

Introduction

Public emergencies occur with a certain frequency (Ni et al., 2023), necessitating timely public opinion analysis and guidance by relevant authorities to effectively respond to incidents that may have significant societal impact (Song et al., 2013; Cai and Golay, 2023). The development of information technology has enabled the rapid acquisition of public opinion data (Scuotto et al., 2017; Mostafa, 2021; Han, 2023), providing a valuable source of emergency information and offering significant support for public opinion analysis and guidance (Gao et al., 2021; Ren et al., 2023). Nonetheless, there exist phenomena whereby individuals and groups attempt to manipulate (Li et al., 2023; Sun et al., 2023) public opinion to maximize personal or specific group interests. The public is susceptible to manipulation, and it is sometimes difficult to determine which opinions are true and objective. When public opinion is subject to manipulation, the objectivity with which the public perceives the state of events can be seriously impeded. In such cases, extreme opinions are amplified and negative emotions spread more rapidly. These phenomena significantly challenge effective public opinion analysis and governance, posing considerable risks to governmental emergency response efforts. Existing studies on public opinion analysis and management have primarily focused on the dynamics of opinion dissemination (Zhao et al., 2022; Hong et al., 2023) and topic extraction (Hui, 2022; Liu et al., 2022). However, the quality of public opinion deserves equal attention, as it reflects the public's capacity to perceive and respond to crisis events. The concept of consensus level in Group Decision Making (GDM) theory provides a suitable framework for evaluating opinion quality. Enhancing consensus among the public can lead to improved opinion quality, thereby strengthening collective perception and response capabilities. Researchers have widely applied GDM perspectives to public opinion studies, particularly in the areas of fuzzy information processing (Zhao et al., 2021; Fu et al., 2022; Liu et al., 2022), uncertain behavior modeling (Li et al., 2023), and consensus measurement (Yang et al., 2024). And this work respond to the problem of public opinion fusion in emergency scenarios through a group consensus reaching method based on the regulation of manipulative behaviors, aiming to obtain high-quality public opinions and improve public perception.

Manipulative behavior may come from individuals or small groups with the aim of obtaining a desired decision result. In recent years, researchers have begun to focus on how to prevent and control manipulative behaviors in group consensus decision-making, and have proposed many models and methods with some achievements. Wu (2023) develops an optimal feedback model that prevents individual and group manipulation by minimizing group adjustment costs and ensures that groups are both efficient and fair in the consensus reaching process (CRP). Sun (2023) proposed a method based on minimum adjustment and maximum entropy to prevent weight manipulation behavior in social network group decision making through power index and feedback mechanism. Dong (2021) proposes a trust relationship manipulation model, analyzes how individuals can influence group decisions by manipulating trust relationships, and explores management methods based on group strategies. Chen (2023) proposes an opinion evolution model based on individual values and constructs an optimized consensus manipulation model that achieves the desired consensus outcome by adjusting the initial opinions while considering the effects of attenuation factors and elasticity of substitution on individual values. Liu (2024) proposed an anti-manipulation weight function based on the uninorm operator for managing subgroup manipulation behavior. Sasaki (2023) explored manipulative behavior in groups from a game-theoretic perspective. In addition, manipulative tendencies are behavioral tendencies of individuals to unduly influence the overall outcome of a decision in a group decision-making process by expressing extreme opinions and exploiting trust relationships.

With the development of machine intelligence (Bolton et al., 2018), many algorithms and intelligences are used to solve group consensus decision making problems. Human-machine collaborative consensus reaching mechanisms can significantly improve decision-making efficiency (Lettieri et al., 2023). The assistance of machine intelligence can efficiently handle complex optimization solution problems. He (2023) proposed a dynamic opinion maximization framework based on reinforcement learning for selecting seed nodes and propagating positive opinions in signed social networks to achieve optimal node opinion distribution. Rui (2022) proposed an adaptive algorithm based on reinforcement learning for dynamically adjusting preference matrices in group decision making to improve consistency and consensus among decision makers. Wang (2022) designed a deep reinforcement learning-based consensus reaching strategy for opinion dynamics to minimize the adjustment cost to enhance group consensus and guide opinions under time constraints. Hassani (2023) proposed a reinforcement learning mechanism based on a deep deterministic policy gradient (DDPG) algorithm for dynamically adjusting feedback parameters and decision maker weights to accelerate the consensus reaching process in group decision making. And for behavior management in GDM, the objectivity of machine intelligence is able to take greater advantage (Hou et al., 2025).

Based on the above analysis, the deficiencies in existing research addressing behavioral management in group consensus decision-making are summarized as follows.

(1) Most related research has been carried out from the perspective of the management of manipulative behavior in groups through weighting adjustment (Wu et al., 2021; Xiong et al., 2024; Liu et al., 2024) or opinion feedback recommendation.(Dong et al., 2021; Gong et al., 2024) While these studies have made important contributions, it is worth noting that group decision-making inherently operates within social network environments. As such, the management of manipulative behavior through adjustments in trust relationships appears to be an essential yet relatively underaddressed dimension. Moreover, although existing studies on the identification of manipulation behaviors have primarily concentrated on extreme opinions and trust dynamics, the potential role of forecasted influence in trust relationships remains insufficiently explored. Introducing iterative processes to estimate expected influence may present meaningful opportunities to further improve the identification of manipulation behaviors.

(2) The integration of machine intelligence into group decision-making has primarily aimed to enhance objectivity in recommending feedback or optimizing the weights of individuals or subgroups, thereby leveraging the computational advantages of intelligent systems. However, relatively limited attention has been given to employing machine intelligence for the management of manipulative behaviors.

As an attempt to complement the limitations, an effective human-machine collaborative dynamic group consensus reaching mechanism that considers manipulative behavior is developed in this work. The main contributions of this research include the following.

(1) A new human-machine collaborative group consensus reaching mechanism is proposed. This mechanism is

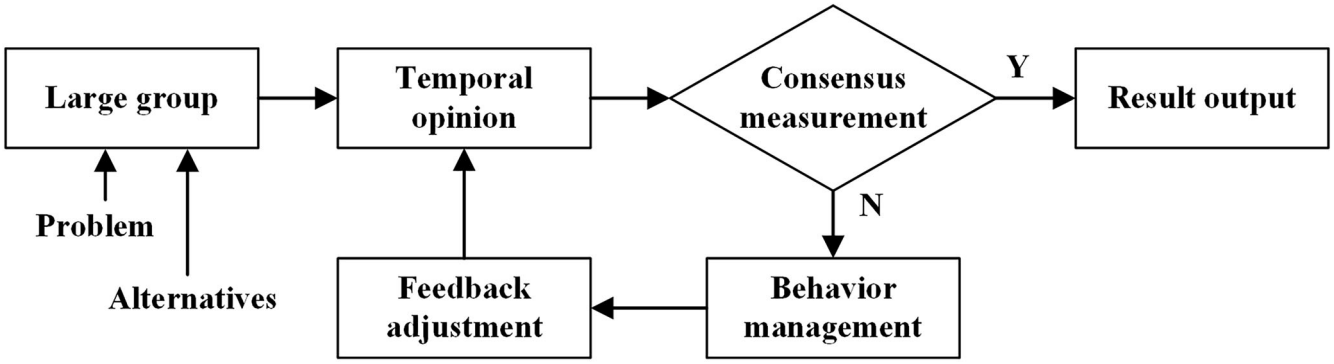


Fig. 1 Traditional behavior-based CRP paradigm.

applicable to behavior management problems, in which the human group expresses its opinions on the problems to be solved and communicates continuously to obtain the collective opinion. The machine moderator learns from the behavioral data generated during the communication within the human group and adjusts the parameters with feedback to control the manipulative behaviors.

(2) A function is proposed to evaluate the degree of manipulation tendency by jointly considering both the extremity of opinions and the anticipated influence within the decision-making process. Manipulation tendency is classified into high, medium, and low levels based on the parameters of the neural element in the uninorm operator and by mapping two-dimensional behavior into degrees of manipulation tendency. This can quantitatively describe the possible degree of manipulation tendency of individuals.

(3) A method for public opinion fusion in emergency response environments is proposed. This method not only improves the consensus level of collective opinions through coevolution but also enhances the reliability of collective opinions through reinforcement learning (RL). This provides some support for analyzing and regulating public opinion in emergency scenarios and improving the accuracy of public perception.

The remainder of this paper is organized as follows. "Preliminaries" introduces some preliminaries to help understand the proposed model. "Human decision-making with manipulative behavior" then introduces the opinion-trust coevolution process and manipulation tendency measurement function in the human decision environment. "Machine moderator training for behavior regulation" reports the training of the machine moderator through RL to feed back the adjustment parameters of the social network, based on which the human-machine collaborative mechanism is summarized. "Case example: Public opinion fusion" presents an application of the proposed mechanism to the fusion of public opinions in the Fujian rainstorm emergency response event. "Discussions" presents some simulation experiments, comparative analysis, limitations and future work. Finally, the conclusions of this research are summarized in "Conclusion".

Preliminaries

This section details some preliminary knowledge helpful for understanding the methodology. These details are related to the behavior-based consensus reaching process (CRP), human-machine collaboration, trust-based social networks, and RL.

Behavior-based CRP. In the context of GDM, evaluating the accuracy of the final decision result often presents a challenge. Researchers have been devoted to enhancing the degree of agreement among participants in large-scale groups, aiming to reach decisions that reflect higher levels of consensus. This process is commonly referred to as the consensus-reaching process (CRP). Compared to conventional GDM scenarios, large group decision-making (LGDM) typically entails the involvement of more individuals, often representing diverse and sometimes conflicting interests, which can lead to more equitable and representative decision results. Let a group of individuals $E = \{e_1, e_2, e_3, \cdots, e_M\} (M \geq 20)$ be the large group. The opinion of individual e_i is o^i, where $o^i \in [0, 1]$. Before reaching a consensus, the members of the decision-making group must go through several rounds of discussion, and it is assumed that consensus is reached at round T. After each round of discussion, the opinion of each individual may change, and the opinion of individual e_i in round t is denoted as o_i^t, $t = 1, 2, ..., T$. Then, the evaluation vector of the large group E in round t is obtained as $O^t = (o_i^t)_{1\times M}$.

As the number of individuals involved in decision-making increases, certain collective behaviors may exert a noticeable influence on the final decisions. Consequently, numerous researchers have explored various behavioral patterns exhibited by decision-makers to foster greater consensus. These include non-cooperative behavior, manipulative behavior, and limited compromise behavior, among others. In the LGDM process, the traditional behavior-based CRP paradigm mainly consists of five components: opinion expression, consensus measurement, behavior management, feedback adjustment and result output. A schematic representation of the traditional behavior-based CRP paradigm[40] is illustrated in Fig. 1.

Trust-based social networks. Social network analysis is a widely used approach for examining interactions among social entities (Wu et al., 2015), offering a clear and visual means of representing both interpersonal trust and social ties. For a group of individuals $E = \{e_1, e_2, e_3, \ldots, e_M\}$, the collection of trust relationships can be denoted as L, while their corresponding social network structure $G = (E, L, V)$ is typically modeled using an adjacency matrix $V = (v_{ij})_{M\times M}$. If $v_{ij} \neq 0$, it indicates that a trust relationship (Liu et al., 2019) exists between individuals e_i and e_j, with the corresponding trust strength represented by $v_{ij} \in [0, 1]$. Conversely, $v_{ij} = 0$ signifies the absence of trust between e_i and e_j. Let the social network of the human group at discussion round t be denoted as G^t, with its structure described by the adjacency matrix $V^t = (v_{ij}^t)_{M\times M}$. Table 1 provides an illustrative example of such a social network, where e_1 trusts e_3, e_2 trusts e_1 and e_3, and e_3 trusts e_1.

Human-machine collaboration. Human-machine collaborative decision-making refers to a cooperative process in which both

humans and intelligent computer systems jointly participate to complete decision-related tasks. Prior research consistently indicates that in certain domains, particularly those characterized by well-structured and complete datasets, machines can produce more optimal decisions than humans (Garai et al., 2023). However, a pivotal study by Wilson and Daugherty (2018) emphasizes that the most substantial performance gains are realized when humans and machines work together. Consequently, a growing consensus supports the adoption of a hybrid decision-making framework, wherein both entities contribute collaboratively. One perspective suggests that machines can directly participate in decision-making alongside human members (Haesevoets et al., 2021), though the extent to which their outcomes are accepted often depends on situational factors, as illustrated in Fig. 2a. Alternatively, some researchers emphasize leveraging machine intelligence to monitor and interpret human behaviors, thereby enhancing decision reliability through behavioral regulation mechanisms (Xiong et al., 2023; Xue et al., 2024), as depicted in Fig. 2b. In order to prevent and control human manipulation due to related interests, the inclusion of objective machine moderators ensures fairness by analyzing human behavioral data for neutral decision regulation. This human-machine collaborative group consensus reaching mechanism can ensure an objective and rational decision result.

Reinforcement learning. RL is a branch of machine learning in which agents attain objectives through continuous interaction with their surrounding environment.(Mnih et al., 2015) In this paradigm, an agent interacts with an environment, taking actions based on its state. The environment, in turn, provides feedback in the form of rewards and the next state. The objective of the agent is to maximize long-term rewards through trial-and-error learning. The fundamental concept of RL revolves around learning from rewards. Agents explore their environment, attempting different actions and adjusting their actions based on the received rewards to optimize future outcomes. This learning methodology differs from supervised and unsupervised learning, as it focuses on maximizing cumulative rewards over extended periods rather than just predicting labels or uncovering patterns in data. The interaction in RL is shown in Fig. 3.

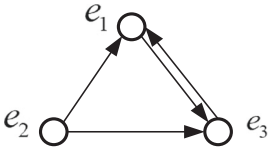
Specifically, Deep Deterministic Policy Gradient (DDPG) (Duan et al., 2020) is an RL algorithm specifically designed for solving problems with continuous action spaces. By integrating deep neural networks with deterministic policy gradient techniques, DDPG employs two neural networks, the actor and the critic, to learn policy and value functions separately, aiming to optimize long-term rewards.

Human decision-making with manipulative behavior

This section first introduces the opinion-trust coevolution mechanism to simulate the process of human groups engaging in a discussion to arrive at a decision result. Then, a function of the degree of manipulation tendency is constructed based on the uninorm operator. Finally, the impact of malicious manipulative behaviors on human decision-making in the absence of regulation is explored via a simulation experiment.

Opinion-trust coevolution. Given the presence of social networks, interpersonal communication becomes inevitable, rendering an individual not entirely independent in their judgments. Consequently, an individual's opinions may be shaped and altered through mutual influence during interactions. Moreover, as discussions unfold, it is not only opinions that undergo changes; trust relationships among participants may also shift in parallel. In certain scenarios, these two dynamics, opinion evolution and trust adjustment, are interdependent and may exert reciprocal effects. This section presents a summary of the CRP under the framework of opinion-trust coevolution, synthesizing insights from existing scholarly research.

Table 1 Social network representation example.

Adjacency matrix	Graph
$V^0 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, $v_{ij} \in [0,1]$	

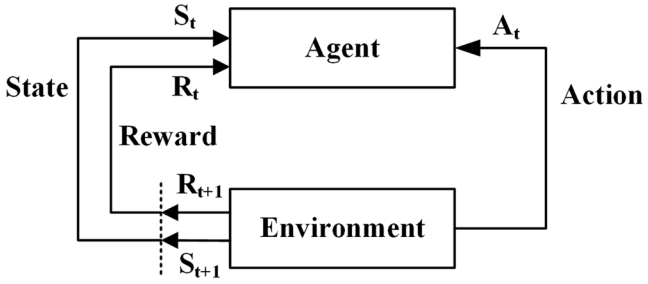


Fig. 3 The interaction in reinforcement learning.

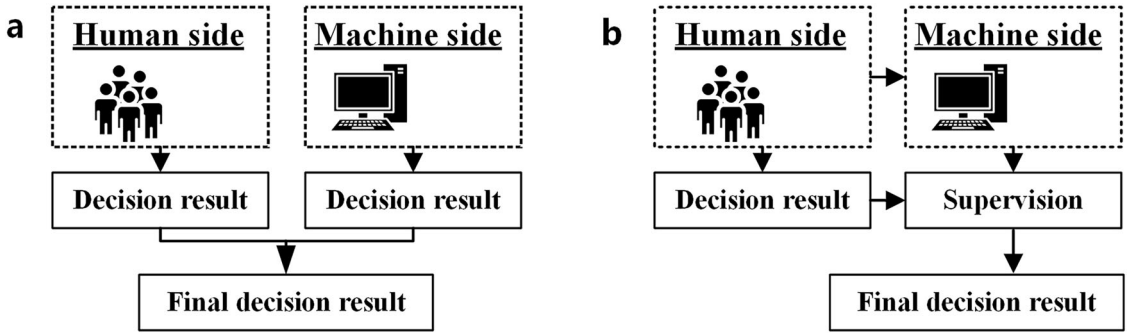


Fig. 2 The two modes of human-machine collaboration. a Collaborative model with simultaneous human and machine decision-making; **b** Collaborative model with human decision-making and machine supervision.

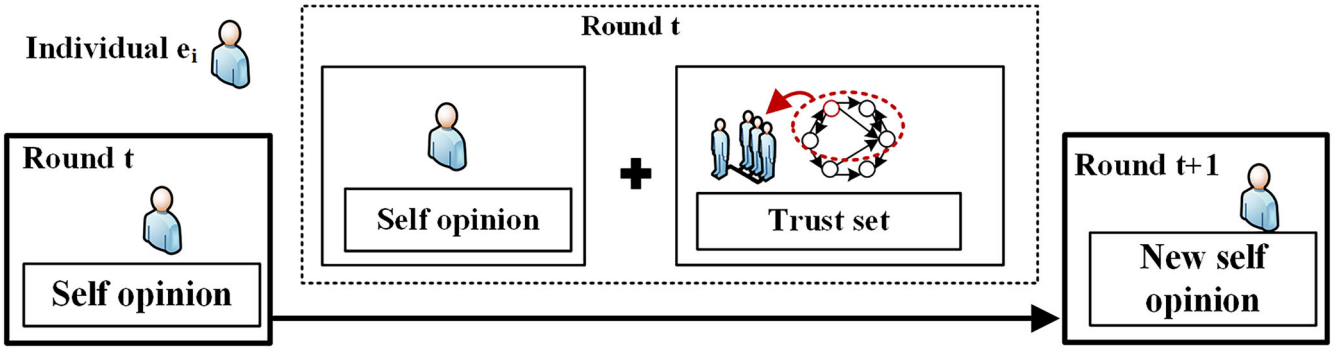


Fig. 4 The opinion update process.

(1) **Opinion updating**

In real-world scenarios, the evolution of individual opinions is shaped by two primary factors. First, individuals differ in their levels of attachment or resistance to changing their initial opinions, a phenomenon often described as opinion stubbornness. Second, the progression of opinion adjustment and consensus formation is inherently influenced by social network structures, whereby individuals are likely to consider the perspectives or evaluative feedback from those they trust.

The opinion of individual e_i in round t is denoted as o_i^t. Simultaneously, the trust-based input information gathered from the social network during the same round is expressed as $trust_i^t = \sum_{j=1}^{M} w_{ij}^t o_i^t$, where

$$w_{ij}^t = \begin{cases} \frac{v_{ij}^t}{\sum_{j=1}^{M} v_{ij}^t}, & v_{ij}^t \neq 0 \\ 0, & v_{ij}^t = 0 \end{cases}.$$

Integrating both personal persistence and external influences from the social network, the opinion of individual e_i is updated in discussion round $t+1$(Jia et al., 2015):

$$o_i^{t+1} = \dot{x}_i * o_i^t + \dot{y}_i * trust_i^t \quad (1)$$

where $\dot{x}_i$ represents the degree of stubbornness individual e_i exhibits toward his/her initial opinion, and $\dot{y}_i = 1 - \dot{x}_i$ denotes his/her receptiveness to information received from trusted connections within the social network. The attitude parameter $\dot{x}_i$, $\dot{y}_i$ captures individual behavioral tendencies, reflecting how each person weighs different sources of opinion. Notably, individuals may assign varying levels of importance to these sources, leading to heterogeneous behavior patterns during opinion updating. The opinion evolution process can be reformulated in matrix notation as follows:

$$O^{t+1} = \dot{X} * O^t + \dot{Y} * Trust^t,$$

where $\dot{X}$, $\dot{Y}$ is the attitude parameter matrix of human group E. The opinion update process of individual e_i is shown in Fig. 4.

(2) **Trust updating**

As individuals continue to exchange opinions and preferences within the group, the underlying trust relationships among them may dynamically evolve. Social judgment theory suggests that individuals are more receptive to opinions that are close to their own and reject those that deviate significantly (Sherif and Hovland, 1961), which explains why trust between individuals tends to strengthen with increasing opinion similarity and decline with divergence. New trust connections are likely to form between individuals who share similar opinions, while existing ties may weaken or even dissolve when significant opinion differences arise.

The distance matrix of human group $E = \{e_1, e_2, e_3, \ldots, e_M\}$ in discussion round t is denoted as $D^t = (d_{ij}^t)_{M\times M}$, where $d_{ij}^t = \sqrt{(o_i^t - o_j^t)^2}$ is the distance between the opinions of individuals e_i and e_j.

The social network trust relationship updating rules based on the group distance matrix D^t are as follows:

$$\bar{v}_{ij}^{t+1} = \begin{cases} \min\left\{1, v_{ij}^t + (1 - d_{ij}^t) * v_{i\,\min}^t\right\}, & d_{ij}^t < r \\ \max\left\{0, v_{ij}^t - v_{i\,\min}^t * d_{ij}^t\right\}, & d_{ij}^t > s \end{cases}, \quad (2)$$

where $r < s$, r is the trust relationship enhanced threshold, s is the trust relationship broken threshold, $v_{i\,\min}^t = \min\{v_{ij}^t | j = 1, 2, \ldots, M\}$, and

$$v_{ij}^{t+1} = \begin{cases} \min\left\{\bar{v}_{ij}^{t+1} + \eta_i * v_{i\,\min}^t, 1\right\}, & \eta_i \geq 0 \\ \max\left\{0, \bar{v}_{ij}^{t+1} + \eta_i * v_{i\,\min}^t\right\}, & \eta_i < 0 \end{cases}. \quad (3)$$

In particular, η_i is the influence adjustment degree parameter for individual e_i; if $\eta_i = 0$, there is no external force intervening in the updating of the trust relationship; this is used in this work to regulate manipulative behavior. The trust relationship update process of individual e_i is shown in Fig. 5.

(3) **Exit rules**

The opinion-trust coevolution process terminates under two conditions: (i) when a consensus is successfully achieved, triggering an automatic exit mechanism; or (ii) when the predefined maximum number of discussion rounds has been completed.

1) Consensus measurement

Euclidean distance is adopted as a measurement metric to quantify the divergence between individual and collective opinions. The distance between the opinion of individual e_i and the collective opinion in round t is defined as

$$d_i^t = \sqrt{(o_i^t - \bar{o}^t)^2}, \quad (4)$$

where $\bar{o}^t = \frac{1}{M}\sum_{i=1}^{M} o_i^t$. The consensus level of individual e_i in round t is defined as follows.

$$CL_i^t = 1 - d_i^t \quad (5)$$

The consensus level of group E in round t is defined as follows.

$$GCL^t = 1 - \frac{1}{M}\sum_{i=1}^{M} d_i^t \quad (6)$$

2) Stable state identification rules

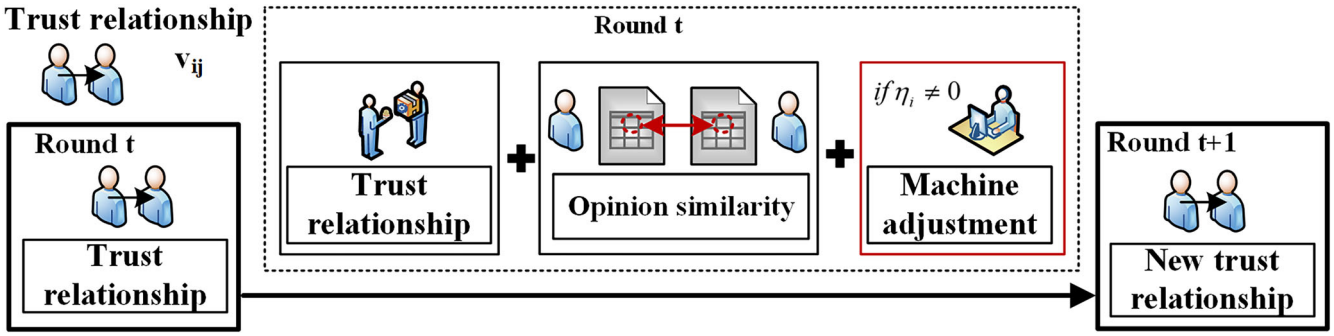


Fig. 5 The trust relationship update process.

The consensus level threshold is denoted as φ and the maximum number of discussion rounds is denoted as $T_{\max}$. The consensus threshold was set to 0.8, in line with prior research in emergency decision-making contexts where rapid agreement is essential (Liu et al., 2022). This threshold reflects a balance between sufficient group agreement and operational efficiency. Higher thresholds may be considered in future research for high-risk decisions requiring stronger unanimity.

The following rules are defined as the termination conditions under which the coevolution is considered to have reached a stable equilibrium, after which the evolution is halted and group opinion $O^{'} = O^t$ is obtained:

i) $GCL^t > \varphi$;

ii) $t \geq T_{\max}$.

Manipulative behavior. It is difficult to judge from the decision result or the level of group consensus whether manipulative behavior exists and whether it has been successful. Therefore, the starting point of judgment must be the two possible ways of implementing manipulative behaviors, i.e., expressing extreme opinions and influencing others. A two-dimensional manipulative behavior tendency measurement model for the opinion expression behavior and influence behavior (i.e., the behavior of being trusted by other individuals) is constructed for individual e_i. First, the extreme opinion degree is proposed to measure human opinion expression behavior; then, the influence degree forecast is proposed to measure influence behavior. On this basis, a function of the degree of manipulation tendency is constructed.

(1) **Extreme opinion degree**

Engaging in manipulative behavior to pursue personal interests results in different degrees of damage to collective interests. Considering the intentionality of manipulative behavior, extreme opinions are the most convenient and effective way to successfully achieve manipulation. Therefore, extreme opinions are significant to the existence of manipulative behavior.

Definition 1. The extreme opinion degree for individual e_i in discussion round t is defined as follows.

$$ife_i^t = \begin{cases} o_i^t, & o_i^t \geq \bar{o}^t + 3 \cdot \sigma(o^t) \\ 0, & \bar{o}^t - 3 \cdot \sigma(o^t) < o_i^t < \bar{o}^t + 3 \cdot \sigma(o^t) \\ 1 - o_i^t, & o_i^t \leq \bar{o}^t - 3 \cdot \sigma(o^t) \end{cases} \quad (7)$$

In the field of modern quality management, Six Sigma theory (Rathi et al., 2017) is widely used as a systematic statistical method for process improvement and defect control. Through the statistical tools and methods of Six Sigma, collective data can be systematically analyzed to effectively identify the relative extreme opinions that significantly deviate from the goal. Opinions outside the interval $[\bar{o}^t - 3 \cdot \sigma(o^t), \bar{o}^t + 3 \cdot \sigma(o^t)]$ are considered to be extreme, while those inside the interval are considered to be normal opinion. Specifically, $\bar{o}^t$ is collective opinion and $\sigma(o^t)$ is the standard deviation of the group opinion. When $ife_i^t = 1$, it is assumed that individual e_i has extreme preferences, which implies that individual e_i is completely confident regardless of the presence of manipulative behavior. If $ife_i^t = 0$, it is assumed that individual e_i has no extreme preferences.

(2) **Influence degree forecast**

If the influence of an individual with an extreme opinion is high, the individual's damage to the collective good is usually greater than that of an individual with less influence. Influence degree refers to the degree to which an individual is capable of affecting others within a group. It is typically quantified by summing the trust intensities that all other members of the group assign to individual e_i. Clearly, the greater the level of trust placed in individual e_i by others, the higher their overall influence within the group.

However, the degree of influence caused by an individual opinion is an a posteriori fact that cannot be measured at the time the opinion is expressed. Thus, the influence of individual opinions must be predicted. The predicted influence degree denotes the potential degree to which an individual e_i may influence others in the group at a specific round t. Unlike static influence, predicted influence degree incorporates temporal dynamics (i.e., the number of discussion rounds) as well as the pairwise opinion similarity between individuals. In group decision-making environments, individuals gradually enhance their understanding of the decision issue through iterative discussions. As a result, their confidence in their own opinions tends to increase over time, while their susceptibility to being influenced by others tends to decrease. As the number of discussion rounds increases, the level of trust between individuals may show an upward trend; however, the influence of trust on opinion formation may gradually weaken. To account for this phenomenon, a time decay function (Maruyama and Moriya, 2021) is introduced into the computation of predicted influence degree. Furthermore, it is evident that the greater the opinion similarity between individual e_i and individual e_j, the more likely it is that e_i will exert an predicted influence degree on e_j. In summary, the measurement of the predicted influence degree is primarily determined by three factors: trust intensity, opinion similarity, and the degree of temporal decay.

Definition 2. The predicted influence degree for individual e_i in discussion round t is defined as

$$imp_i^t = g(t) \cdot \sum_{j=1, j \neq i}^{P} \left[tr_{ji} \cdot \left(1 - \left|o_i^t - o_j^t\right|\right)^{\beta} \right], \quad (8)$$

where tr_{ji} denotes the degree of trust of individual e_j in

individual e_i, β represents the sensitivity of opinion similarity to trust. In addition, $g(t) = e^{-\alpha(t+l)}$ is a time decay function constructed on the basis of Newton's law of cooling (Maruyama and Moriya, 2021), where α is the attenuation coefficient and the value of $g(t)$ decays exponentially as t increases, meaning that the degree of influence of the opinion diminishes with time. This is in line with reality because as time goes on, an individual's understanding of a problem becomes progressively deeper.

(3) Degree of manipulation tendency

To prevent and control possible malicious manipulation in a group discussion, i.e., a GDM process, the degree of manipulation tendency must be quantified. To do so, the uninorm operator is first introduced. Uninorm aggregation operators, proposed by Yager and Rybalov,[39] serve as a unified framework that encompasses both t-norm and t-conorm operations. The uninorm operator is applicable to the mapping of two-dimensional data; this reflects different enhancement characteristics, both positive and negative, as the input two-dimensional behavioral data changes. Considering the characteristics of manipulative behavior, the degree of manipulative behavior tendency embodied in two-dimensional behavior is quantified based on the uninorm operator.

Definition 3. A uninorm (Yager and Rybalov, 1996) is a mapping expressed as $U : [0,1] \times [0,1] \to [0,1]$, which has the following properties for all $a, b, c, d \in [0,1]$:

i. Commutativity: $U(a,b) = U(b,a)$;
ii. Monotonicity: $U(a,b) \geq U(c,d)$ for $a \geq b, c \geq d$;
iii. Associativity: $U(a, U(b,c)) = U(U(a,b), c)$;
iv. Neutral element: $\exists g \in [0,1]$: $U(a,g) = a$ for all a.

In contrast to t-norms and t-conorms, whose neutral elements are fixed at 1 and 0, respectively, uninorms allow the neutral element to be any value within the unit interval. The definitions of manipulation thresholds can be reflected by the neutral element g. For different decision problems, different thresholds are selected to portray the degree of manipulation tendency. In this work, a form of the uninorm operator proposed by Foder (1997) is adopted, as follows.

$$U(a,b) = \begin{cases} \frac{ab}{g} & , \quad if\, 0 \leq a, b \leq g \\ \frac{a+b-ab-g}{1-g} & , \quad if\, g \leq a, b \leq 1 \\ \frac{a+b}{2} & , \quad else \end{cases} \tag{9}$$

If all the two-dimensional information of the input operator is greater than the value of the neutral element, the two-dimensional information will be positively reinforced for integration; if all the two-dimensional information of the input operator is less than the value of the neutral element, it will be negatively reinforced for integration. The function of the degree of manipulation tendency can now be introduced.

Definition 4. The degree of manipulation tendency MT_i^t of individuals e_i in discussion round t can be calculated by $MT_i^t = U(ife_i^t, imp_i^t)$, which can be expressed as follows.

$$MT_i^t = \begin{cases} \frac{ife_i^t \cdot imp_i^t}{g} & , if\ 0 \leq ife_i^t,\ imp_i^t \leq g \\ \frac{ife_i^t + imp_i^t - ife_i^t \cdot imp_i^t - g}{1-g} & , if\ g \leq ife_i^t,\ imp_i^t \leq 1 \\ \frac{ife_i^t + imp_i^t}{2} & , if\ \min(ife_i^t, imp_i^t) \leq g \leq \max(ife_i^t, imp_i^t) \end{cases} \tag{10}$$

Based on the setting of the neutral element g, which is called the manipulation threshold in this work, via the uninorm operator, the individuals involved in decision-making can be categorized into the following three main groups by measuring the extreme opinion degree and the predicted value of influence.

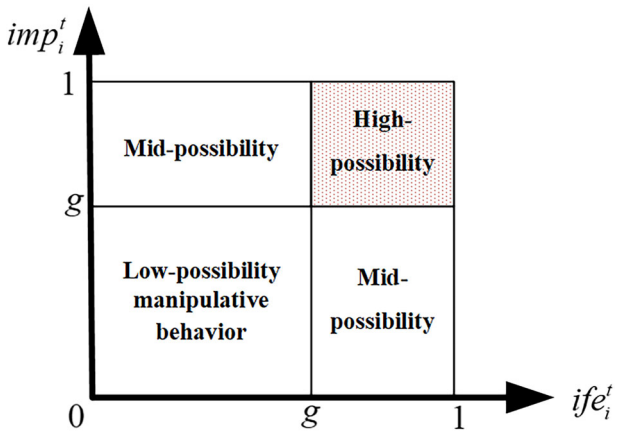


Fig. 6 Behavior classification with manipulation threshold g.

i. High possibility of manipulative behavior: $g \leq ife_i^t, imp_i^t \leq 1$ When the opinion of individual e_i is more extreme and more influential, this individual is considered to have the intention and ability to manipulate the group, and the degree of his/her possible manipulation tendency is positively reinforced by the uninorm operator.
ii. Low possibility of manipulative behavior: $0 \leq ife_i^t, imp_i^t \leq g$ When the opinion of individual e_i is less extreme and his/her influence is weaker, the individual is considered to have neither the intention nor the ability to manipulate the group, and the degree of his/her possible manipulation tendency is reduced by the uninorm operator.
iii. Medium possibility of manipulative behavior: $\min(ife_i^t, imp_i^t) \leq g \leq \max(ife_i^t, imp_i^t)$

When either the influence or the extreme opinion degree of individual e_i exceeds the threshold g, he/she may have a manipulation intention or may have the ability to manipulate the group. For such individuals, no special treatment is needed, and the degree of his/her possible manipulation tendency can be averaged by the uninorm operator.

The graph in Fig. 6 illustrates the classification of manipulative behavior tendency based on the uninorm operator. The extent of the overall manipulation trend of a group can be measured based on the degree of manipulation tendency.

Definition 5. The group degree of manipulation tendency MT^t of human group E in discussion round t can be calculated as follows.

$$MT^t = \frac{1}{M} \sum_{i=1}^{M} MT_i^t \tag{11}$$

Manipulation tendency in human decision-making. A simulation experiment was designed to observe how manipulative behavior affects group discussion, i.e., human decision-making. The proportion of individuals with a high possibility of manipulative behavior was gradually increased during the opinion-trust coevolution process, and the corresponding group degree of manipulation tendency, group consensus degree, number of group discussion rounds, and decision-making results were observed. The experiment was conducted 100 times, and the results were averaged to obtain the final observations. It should be clarified that, in this simulation setting, individuals with high manipulative tendencies are intentionally assigned extreme preferences biased toward 1. The purpose of this design is to analyze

the influence of unidirectional manipulation on collective opinion dynamics.

The initial opinion was generated randomly between 0 and 1, and the initial social network trust relationships among individuals were randomly generated using a Barabási–Albert (BA) scale-free social network. The other parameters were set as follows: the number of individuals $M = 100$, the consensus level threshold (Liu et al., 2024) $\varphi = 0.8$, the stable threshold $T_{\max} = 20$, the manipulation threshold $g = 0.8$, and the attitude parameters x_i, y_i of individual e_i were generated randomly. The independent and observed variables of the experimental design are summarized in Table 2, and the results of the simulation experiment are exhibited in Fig. 7.

The following three conclusions can be drawn from Fig. 7.

1. With the increase of the proportion of individuals with a high possibility of manipulation tendency high-possibility manipulation tendency proportion, both the initial and final group manipulation tendency increased, with the increase of final group manipulation tendency being greater. This may be because, during the group discussion process, a manipulator continues to convert new manipulators to motivate the collective opinion to reach the manipulator's preferred opinion.
2. From a consensus perspective, the difference in the final group consensus level was not large, but the manipulative behavior caused an increase in the number of discussion rounds required to reach a consensus. However, when the high-possibility manipulation tendency proportion was too high, the number of discussions did not continue to increase, which is reasonable; when most of the group members hold extreme opinions and are highly influential, the correct decision result may be an extreme value.
3. The final collective opinion differed greatly from the correct opinion. The initial opinion was randomly generated between 0 and 1, which indicates that the final collective opinion comes to about 0.5 is the correct decision-making results. Observation of the purple dotted line in Fig. 7b shows that as the proportion of high manipulation trendsetters increases, the collective opinion is significantly biased from 0,5 to 1. Therefore, the existence of manipulative behavior easily pushes the collective opinion toward the manipulator's preferred opinion.

Table 2 Variable descriptions.

Variable	Symbolic
High-possibility manipulation tendency proportion	P_{HM}
Initial group manipulation tendency	$MT_{initial}$
Final group manipulation tendency	MT_{final}
Group consensus level	GCL_{final}
Discussion rounds	T_{final}
Collective opinion	O'

Machine moderator training for behavior regulation

The preceding content indicates that unregulated opinions and the mechanisms governing social network evolution are susceptible to deliberate manipulation, which consequently leads to diminished reliability of collective opinions derived from information fusion. Hence, with the goal of elevating the consensus level within collective group opinions, the utilization of RL methodologies is proposed to introduce machine moderators for the supervision of manipulative behaviors via the adjustment of group trust relationships.

In this section, a Markov process for coevolution is first constructed, and the machine moderator agent is trained by the DDPG algorithm. Finally, the proposed human-machine collaborative decision-making mechanism that considers manipulative behaviors is summarized.

The construction of a Markov decision process. The proposed reinforcement learning model is built upon the following foundational assumptions: (1) Markov property: the opinion–trust coevolution follows a Markov process, enabling decision-making based solely on the current state; (2) continuity of spaces: both state (group consensus) and action (influence adjustment) are continuous, aligning with DDPG's requirements for high-dimensional control; (3) reward design: the reward function integrates consensus level and manipulation tendency, guiding the agent to enhance consensus while reducing manipulation; (4) behavioral response: individuals update opinions iteratively, and machine moderation effectively reshapes trust ties, ensuring agent influence; (5) scalability: the model generalizes to varied group sizes and network structures, provided dimensional consistency is maintained. These assumptions underpin the model's theoretical soundness and practical applicability in regulating manipulative behavior in group decision-making.

Then, given that past studies have confirmed the Markov properties of opinion-trust coevolution processes, the Markovian environment can be constructed (Hassani et al., 2023). The

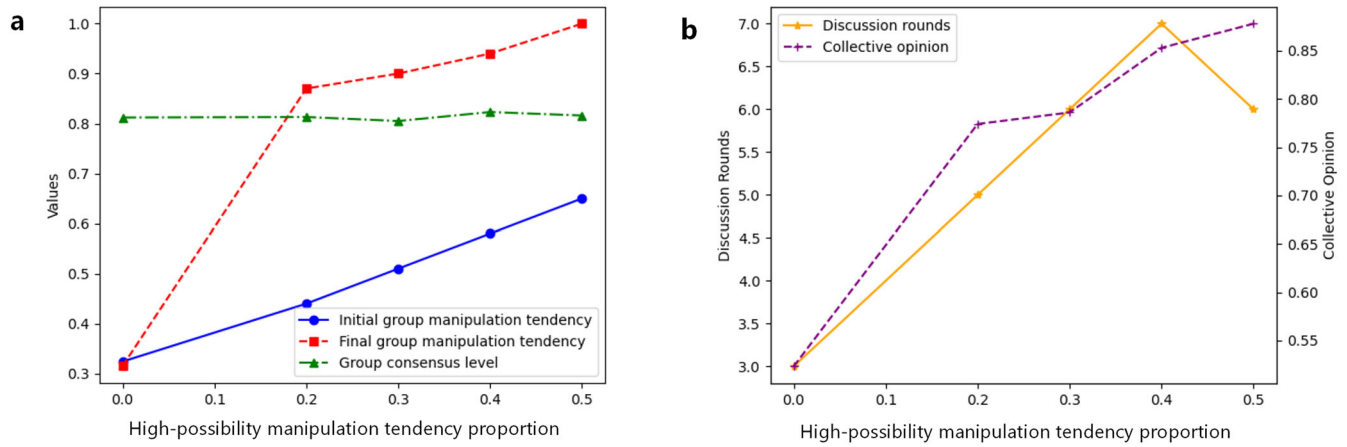


Fig. 7 Simulation result of manipulation tendency. a $MT_{initial}$, MT_{final} and GCL_{final} with different high-possibility manipulation tendency proportion; **b** T_{final} and O' with different high-possibility manipulation tendency proportion.

Markov decision process encompasses three primary elements, namely the state space, action space, and reward function. For the problem of manipulative behavior regulation in public opinion management, these three elements are constructed as follows.

(1) State space
The state space is set to encompass all potential individual opinions that may arise. Suppose that M individuals participate in the group discussion; the state space is represented as $S = \{s_i | s_i \in [0,1], i = 1,2,\ldots,M\}$, where s_i is the opinion of individual e_i, i.e., $s_i = o_i$.

(2) Action space
The adjustment parameters of individuals' influence recommended by the machine moderator, which are also the parameters for adjusting social network trust relationships, are considered the action space $A = \{a_i | a_i \in [-1,1], i = 1,2,\ldots,M\}$, where a_i denotes the influence adjustment degree for individual e_i, i.e., $a_i = \eta_i$. In this context, when $a_i > 0$, the influence of individual e_i is increased to its maximum extent (denoted as 1) by recommending the opinions of this individual to more people on public platforms. Conversely, when $a_i < 0$, the influence of individual e_i is decreased to its minimum extent (denoted as -1) by blocking the opinions of this individual from reaching more people on public platforms.

(3) Reward function

The reward function is used to provide feedback on the actions of the agent, thus guiding the agent to take better actions. For the manipulative behavior regulation problem, the reward function consists of the group consensus level and the group manipulation tendency, and is calculated as follows:

$$rr^t = \omega \cdot rr_1^t + (1-\omega) \cdot rr_2^t, \tag{12}$$

where ω is the coefficient between rewards rr_1^t and rr_2^t. One the one hand, for emergency problems, it is necessary to quickly obtain public opinions with high consensus levels. Thus, the group consensus level must be set as part of the reward function rr_1^t, which is calculated as follows.

$$rr_1^t = GCL^t \tag{13}$$

On the other hand, collective opinions under the influence of manipulative behavior are biased and less reliable. Thus, the group degree of manipulation tendency is used as a partial reward function rr_2^t, which is calculated as follows.

$$rr_2^t = -MT^t \tag{14}$$

Training of the η-agent. According to the constructed Markov decision process, η-agent training is further proposed based on the DDPG algorithm, which is attributed to the continuous nature of both the environment state (the consensus level of individuals) and the agent's actions (the values for all η_i). DDPG agents leverage two distinct networks respectively known as the "actor" and the "critic." The actor network $\tau(s|\theta^\tau)$, whose input is environment states, exploits an action, while the critic network $Q(s,a|\theta^Q)$ makes use of the environment states and the generated action by the actor network to estimate the expected reward.

The η-agent training process necessitates the implementation of a replay buffer Φ. This component stores critical interaction tuples $\Phi(s_1, rr, s_F)$, with s_1 representing the environmental state at the beginning of an episode, rr denoting the immediate reward obtained by the agent, and s_F corresponding to the terminal state observed after action execution. To optimize the critic network, K randomly selected transition tuples are first extracted from the replay memory. These sampled experiences are then utilized to compute and minimize the subsequently defined loss function:

$$L(\theta^Q) = \frac{1}{K}\sum_i (z_i - Q(s_i, a_i|\theta^Q))^2, \tag{15}$$

where K is the size of the minibatch, and the target $z_i = rr_i + \gamma \cdot Q'(s_{i+1}, \tau'(s_{i+1}|\theta^{\tau'})|\theta^Q)^2$, with γ being the discount factor. z_i is modeled via target critic Q' and actor τ' networks. In the training pipeline, the agent executes synchronized updates to its target network at predefined intervals. Concurrently, the parameters of the policy network undergo optimization through gradient ascent, driven by the maximization of this strategic performance metric:

$$\nabla_{\theta^\tau} J \approx \frac{1}{K}\sum_i \nabla_a Q(s,a|\theta^Q)|s = s_i, a = \tau(s_i)\nabla_\theta^\tau \tau(s|\theta^\tau)|_{s_i}. \tag{16}$$

The training of the η-agent is summarized in Algorithm 1. In each episode, the agent takes an action to adjust η in each step. This is continued until the group reaches the desired consensus level with a low manipulation tendency. The parameters of the RL algorithm in this paper are set as follows: the critic learning rate to $\rho_c = 0.001$, the actor learning rate to $\rho_a = 0.0001$, the discount factor to $\gamma = 0.99$, the size minibatch to $K = 128$, and the total number of episodes to $N = 15000$.

Algorithm 1. Training of the η-agent

Input: critic learning rateρ_c, actor learning rateρ_a, discount factor γ, total number of episodes$N$, exploration noise Ψ

Output: η

```
Initialize the replay buffer Φ;
Initialize the critic network Q(s,a|θ^Q) with random weights θ^Q and the actor network τ(s|θ^τ) with random weights θ^τ ;
Initialize the target network Q' and τ' with θ^Q' ← θ^Q and θ^τ' ← θ^τ;
for episode = 1 to N do
    Reset the decision environment and compute the initial observation s_1;
    Select action η = τ(s_1|θ^τ) + Ψ with exploration noise Ψ;
    Execute coevolution process with action η, calculate the reward rr with Eq. (12);
    Compute the next state s_F based on the coevolution process;
    Store tuple (s_1, rr, s_F) in the replay buffer Φ;
        if Φ contains enough samples for a minibatch then
            Sample a minibatch of size K from Φ;
            for each sample in minibatch do
                Compute target Q-value using target networks: z_i = rr_i + γ · Q'(s_{i+1}, τ'(s_{i+1}|θ^τ')|θ^Q)^2;
                Update critic network by minimizing the loss: L(θ^Q) = (1/K) Σ_i (z_i − Q(s_i, a_i|θ^Q))^2;
            end
            Update actor network using gradient ascent: ∇_{θ^τ} J ≈ (1/K) Σ_i ∇_a Q(s,a|θ^Q)|s = s_i, a = τ(s_i) ∇_θ^τ τ(s|θ^τ)|_{s_i};
            Update target networks: critic θ^Q' ← ρ_c θ^Q + (1 − ρ_c) θ^Q' and actor θ^τ' ← ρ_a θ^τ + (1 − ρ_a) θ^τ';
end
```

Human-machine collaborative mechanism. For the decision problem to be solved, a human group is formed by assembling relevant people to express their opinions and social network trust relationships. The consensus level of the human group is improved by communicating repeatedly, and in each round of discussion, the opinions of individuals and the trust relationships between them are updated.

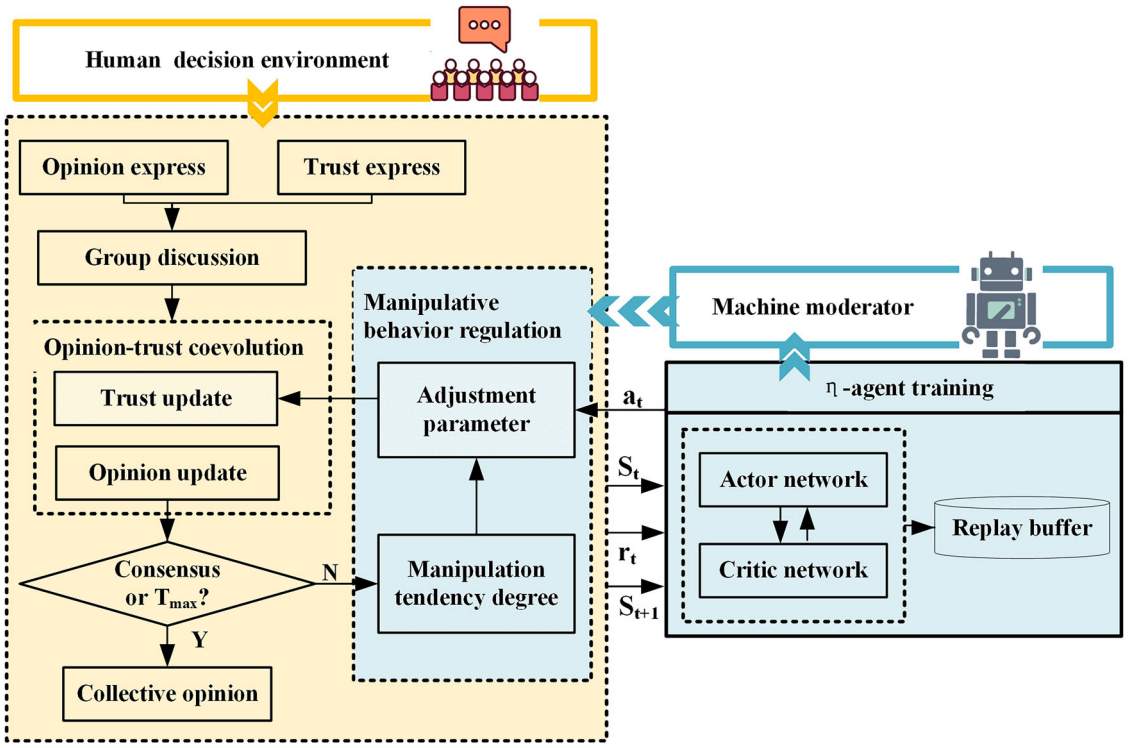


Fig. 8 Human-machine collaborative decision-making mechanism.

To control manipulative behavior while improving the consensus level, an agent is introduced as a machine moderator, with the consensus level and the group degree of manipulation tendency as the objective function. The agent is continuously trained by the DDPG algorithm to recommend the optimal adjustment parameter for updating the social network relationship between individuals. Until the opinion is no longer updated (when the consensus level is sufficient or the time threshold is reached), the collective opinion of the group at that moment is assembled as the output of the decision-making results. The human-machine collaborative consensus reaching mechanism is detailed in Fig. 8.

The specific process of the proposed human-machine collaborative consensus reaching mechanism is summarized as follows.

Step 1. Construction of the initial human decision environment

First, the relevant decision-making group $E = \{e_1, e_2, e_3, \ldots, e_M\}$ is gathered based on the problem to be solved, and each decision-making member expresses his/her initial opinion in the form of a rating between 0 and 1. Then, the initial social network trust relationship of the decision-making group is established based on the expression of the degree of trust of each decision-making member in the form of a number between 0 and 1. Then, $t = 0$.

Step 2. Group discussion

First, $t = t + 1$. Then, the group discussion is initiated, and the updated opinion and the social network trust relationships after the group discussion, respectively denoted as O^t and V^t, are recorded. Group discussion is the process of opinion-trust coevolution, whereby decision-making members respectively update their opinions and their trust relationships based on Eqs. (1) and (2) when no real data is available.

Step 3. Consensus measurement

The group consensus level GCL^t in the current round is measured by Eq. (6). If $GCL^t > \varphi$, the process skips to Step 5. If $GCL^t \le \varphi$, it is determined whether the maximum number of discussion rounds T_{max} has been reached. If $t \ge T_{max}$, then the process skips to Step 5; otherwise, the process continues to Step 4.

Step 4. Manipulative behavior regulation

First, the individual degree of manipulation tendency MT_i^t is measured via Eq. (10). Second, the group degree of manipulation tendency MT^t is measured via Eq. (11). Then, to manage the manipulative behavior and decrease the manipulation tendency while increasing the consensus level, the machine moderator, namely the η-agent, is trained by Algorithm 1 to recommend the trust relationship adjustment parameters. Finally, the process returns to Step 2.

Step 5. Collective opinion fusion

The final collective opinion with a high consensus level is obtained as $O' = O^t$.

Case example: Public opinion fusion

The public opinion fusion case comprises three parts: case background, data processing and result analysis.

Case background. The frequency of natural disasters has increased in recent years, making it particularly important to study responses to disaster emergencies (Lillywhite and Wolbring, 2022; Ashraf et al., 2023). Understanding how to respond effectively to natural disasters, including the formulation of emergency response plans, the rapid mobilization of resources, and the organization of rescue operations, is crucial to safeguarding people's lives and property. Through the study of disaster response (Abounacer et al., 2014; Morgan and Fa Aui, 2018; Ni et al., 2018), it is possible to continuously summarize lessons learned, improve response and public perception capacity, reduce losses, and provide an important reference for building a safer and more stable society. At 08:00 on April 3, 2024, Fujian Province, China, was hit by a torrential rainstorm, resulting in flooding and severe disasters in many areas. In the face of this

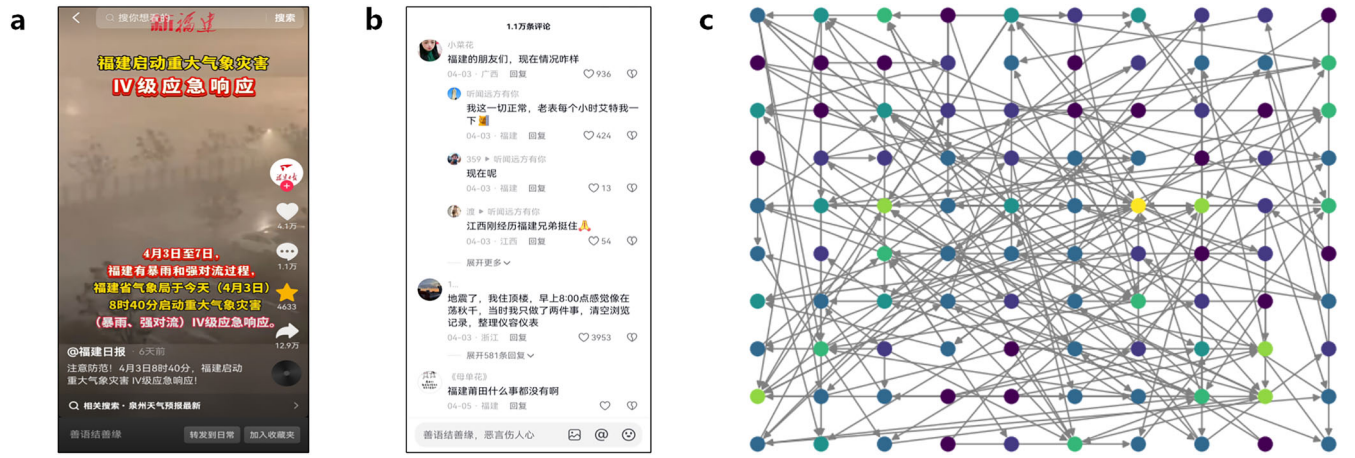


Fig. 9 Related reports, comments, and social network on the case of Fujian rainstorm. a Screenshot of the related report posted by the official "Fujian Daily" account; **b** Screenshot of some of the comments on the related report; **c** Initial social network among related users.

sudden natural disaster, the Fujian Provincial Government quickly activated a Level IV emergency response and released a report through the official "Fujian Daily" account. During the emergency response to the rainstorm, the social media platform TikTok became an important channel for people to express their emotions and concerns about the disaster. Many users posted rainstorm-related comments under the reports, expressing their attitudes toward the emergency event. Related reports and comments are shown in Fig. 9a and 9b.

Data processing. To empirically evaluate the proposed human-machine collaborative consensus reaching mechanism, this study selects public opinion data from the social media platform TikTok during a real-world emergency as the experimental dataset. The full process of data acquisition and preprocessing is outlined as follows.

(1) Data collection
Publicly available user comment data and interaction data (including reposts, likes, and replies) were collected under the official post published by Fujian Daily on TikTok regarding the Level IV emergency response. The data collection window spanned from 08:00 to 18:00 on April 3, 2024, capturing 20 discussion rounds at an hour interval. To ensure analytical validity, we selected the top 100 active users based on interaction frequency.

(2) Opinion and trust construction
To quantitatively represent subjective public opinions extracted from TikTok comment texts, this study applies a lexicon-based sentiment analysis method. First, all collected comments were preprocessed using the Jieba Chinese word segmentation tool, which includes removing stop words and performing tokenization based on Chinese grammar and semantics. Second, the sentiment polarity of each word in the comment was determined with HowNet sentiment lexicon, which classify words into positive (e.g., "安全", "感动", "希望") and negative (e.g., "危险", "担忧", "愤怒"). Then, the sentiment score of user e_i in round t can be calculated by

$$ss_i^t = \sum_{j=0}^{\#wn_i^t} ws_{ij}^t \cdot mf_{ij}^t, \tag{17}$$

where, $\#wn_i^t$ is the number of sentiment words after segmentation for all comments posted by user e_i in round t, ws_{ij}^t is the word strength of the j-th sentiment word, mf_{ij}^t is the modification factor of the jth sentiment word, which mainly includes adverbs of degree and negatives (reversed polarity). Finally, the opinion of user e_i in round t can be quantified as

$$o_i^t = \frac{ss^t - ss_{min}^t}{ss_{max}^t - ss_{min}^t}, \tag{18}$$

where, ss_{max}^t and ss_{min}^t are the maximum and minimum of all the sentiment scores in round t, respectively.
To quantify the subjective trust relationships among users in the online network, this study maps user interaction behaviors, including reposts, likes, and comments, into numerical trust strength values. The interaction strength toward user e_i to user e_j in round t can be calculated by

$$ist_{ij}^t = 0.5 \cdot R_{ij}^t + 0.3 \cdot L_{ij}^t + 0.2 \cdot C_{ij}^t, \tag{19}$$

where, R_{ij}^t, L_{ij}^t and C_{ij}^t are the number of reposts, likes, and comments toward user e_i to user e_j in round t, respectively. Then, trust strength toward user e_i to user e_j in round t can be quantified as

$$v_{ij}^t = \frac{ist_{ij}^t - ist_{min-i}^t}{ist_{max-i}^t - ist_{min-i}^t}, \tag{20}$$

where, ist_{max-i}^t and ist_{min-i}^t are the maximum and minimum of all the interaction strength of user e_i in round t, respectively.
The public opinions without manipulative behavior regulation of the 100 chosen users are reported in Table 3 and the initial social network is shown in Fig. 9c.

(3) Attitude parameter fitting
Based on the constructed data, the attitude parameter of each user was fitted. Figure 10 presents a boxplot of the fitted stubbornness degree parameter $\dot{x}_i$ and a histogram of the cumulative probability distribution. The degree of public stubbornness was mostly distributed around 0.6, which can easily be incited and is applicable to the mechanism proposed in this work.

(4) Manipulation threshold determination

Based on the fitting results of the above user group attitude parameters, we can manage the manipulation behavior through the human-machine collaborative consensus reaching mechanism. The other parameters (Wu et al., 2021) were set as follows: the consensus level threshold $\varphi = 0.8$, the stable threshold $T_{\max} = 20$. However, the parameter setting of the manipulation

Table 3 Unregulated opinions in 20 rounds.

Rounds	e_1	e_2	e_3	e_4	e_5	e_6	...	e_{95}	e_{96}	e_{97}	e_{18}	e_{99}	e_{100}
1	0.74	0.61	0.11	0.98	0.68	0.24	...	0.14	0.74	0.40	1.00	0.71	0.12
2	0.27	0.77	0.67	0.49	0.20	0.00	...	0.70	0.09	0.18	1.00	0.63	0.60
3	0.99	0.97	0.05	0.03	0.14	0.47	...	0.33	0.48	0.75	0.14	0.00	0.97
4	0.13	0.93	0.74	0.99	0.55	0.20	...	0.04	0.90	0.63	0.53	0.34	0.72
5	0.86	0.34	0.97	0.74	0.83	0.80	...	0.86	0.65	0.64	0.87	0.63	0.56
6	0.25	0.66	0.05	0.34	0.95	0.79	...	0.33	0.64	0.65	0.10	0.20	0.55
7	0.42	0.60	0.24	0.59	0.01	0.49	...	0.20	0.20	0.06	0.69	0.08	0.48
8	0.07	0.65	0.73	0.04	0.90	0.79	...	0.28	0.76	0.61	0.05	0.26	0.95
9	0.33	0.13	0.24	0.46	0.43	0.36	...	0.71	0.11	0.84	0.59	0.97	0.07
10	0.16	0.16	0.61	0.22	0.62	0.83	...	0.05	0.27	0.65	0.79	0.54	1.00
11	0.75	0.27	0.73	0.53	0.91	0.23	...	0.99	0.32	0.76	0.77	0.41	0.23
12	0.64	0.88	0.77	0.74	0.29	0.32	...	0.66	0.43	0.55	0.55	0.61	0.27
13	0.80	0.76	0.03	0.09	0.96	0.63	...	0.51	0.00	0.00	0.78	0.54	0.21
14	0.55	0.33	0.94	0.04	0.13	0.58	...	0.45	0.02	0.69	0.77	0.74	0.82
15	0.35	0.77	0.14	0.58	0.32	0.57	...	0.22	0.32	0.46	0.14	0.15	0.13
16	0.54	0.83	0.64	0.29	0.04	0.45	...	0.63	0.86	0.33	0.23	0.98	0.94
17	0.14	0.17	0.55	0.65	0.02	0.33	...	0.39	0.62	0.90	0.22	0.81	0.71
18	0.78	0.99	0.12	0.73	0.25	0.79	...	0.88	0.71	0.03	0.60	0.25	0.65
19	0.09	0.17	0.85	0.21	0.80	0.79	...	0.62	0.14	0.93	0.90	0.41	0.88
20	0.21	0.12	0.42	0.57	0.32	0.29	...	0.69	0.74	0.05	0.65	0.57	0.58

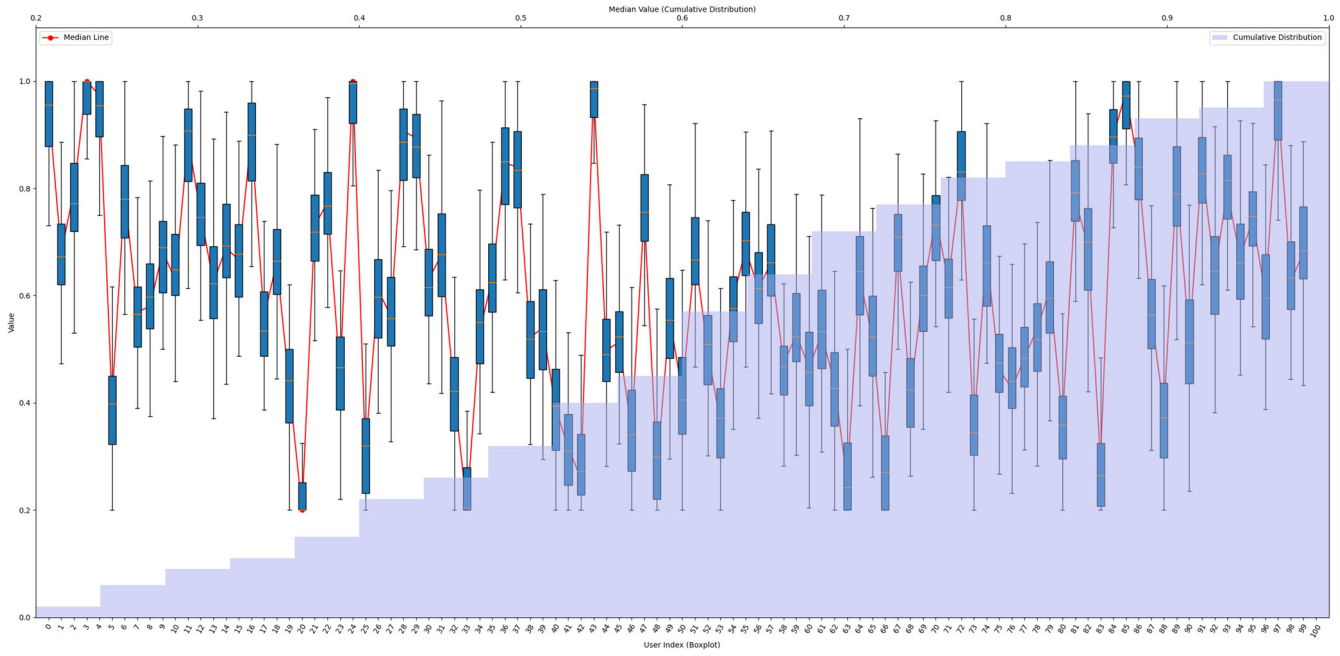


Fig. 10 Distribution the fitted public stubbornness degree. Attitude parameters were fitted to the selected 100 users based on the relevant case data collected.

threshold is difficult to determine subjectively, and we select the most appropriate manipulation threshold through simulation. The speed of consensus reaching under different manipulation thresholds, i.e., the number of iteration rounds needed to reach consensus, is recorded in Table 4. The results show that the manipulation threshold to $g = 0.8$, achieves optimal consensus speed while ensuring stable detection of manipulative tendencies, and was adopted in the case study.

Table 4 The iteration rounds under different manipulation threshold.

Manipulation threshold	0.7	0.75	0.8	0.85	0.9	0.95
Iteration rounds	6	6	5	6	8	10

Result analysis. After the manipulation threshold was determinate, the proposed human-machine collaborative mechanism was used to regulate the manipulative behavior in the chosen public group. The comparative results are exhibited in Fig. 11. On the one hand, as presented in Fig. 11a, the proposed human-machine collaborative decision-making mechanism accelerated the consensus reaching speed. It can be seen that the public opinion reaches a consensus level of 0.8 around 15:30, while the intervention of the human-machine collaborative mechanism allows the public to reach a higher quality of consensus around 13:40. Although the public opinion eventually reached

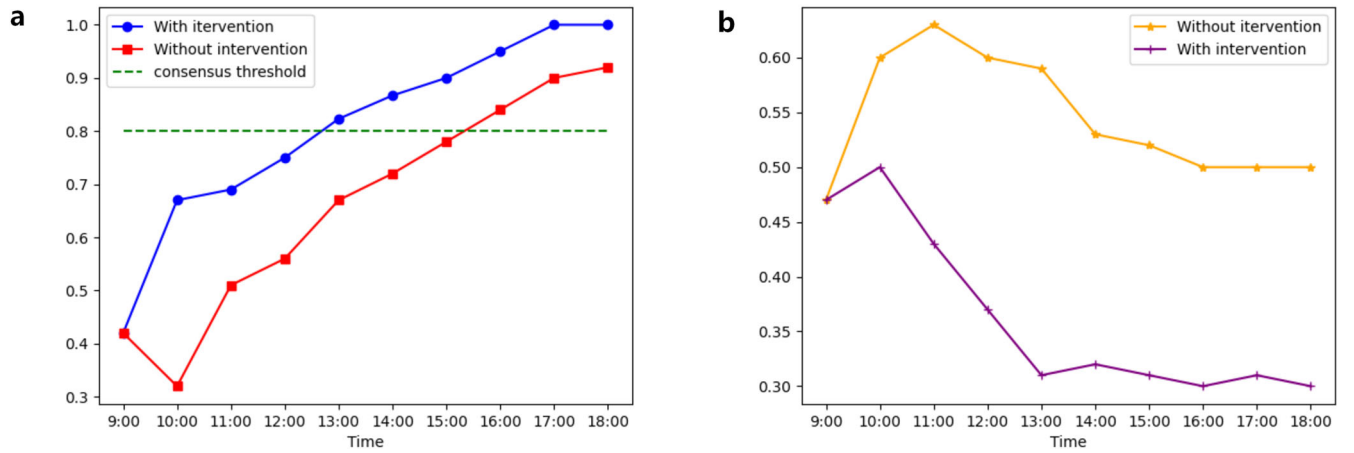


Fig. 11 Comparison results between with and without intervention. a Public consensus level comparison between with and without intervention; **b** Collective public opinion comparison between with and without intervention.

consensus even without the intervention. However, in the Fujian rainstorm event, the public consensus reached almost two hours faster with the human- machine collaborative mechanism intervention, and the faster formation of high quality public's situational awareness may have more preparation time when the public responds to the crisis event, which is important in the emergency situation.

On the other hand, as can be seen from Fig. 11b, before the proposed mechanism intervened, the collective opinion first increased and then decreased to around 0.5. This is far from the value of 0.3 that was reached after the mechanism intervened. The day after the occurrence of the rainstorm event, the official Fujian Daily account reported that the emergency response level was adjusted from Level IV to Level III, indicating that the actual disaster situation was more serious than the previous estimation. In this case, the collective opinion reaching a value of 0.3 is more in line with the actual situation; the public held negative attitudes toward both the rainstorm event itself and the report that the emergency response level was set at Level IV. And the public opinion without intervention has been staying near 0.5, which says that the public concerned about the rainstorm in Fujian is far from enough to perceive the emergency state of affairs crisis. In this paper, the human-machine collaborative mechanism considering the manipulation behavior guides the direction of public opinion better, and recommends the more realistic opinions to the public, which provides a certain reference value for public opinion analysis and public opinion guidance.

The findings indicate that the proposed mechanism can not only quickly integrate high-consensus public opinions, but can also correct the bias that occurs when public opinions are influenced by incitement comments. Public opinions in line with the actual situation are ultimately obtained. Managing manipulative behaviors in public opinion through human- machine collaborative consensus reaching mechanisms can enhance the public's ability to perceive crisis situations, which provides some guidance for public opinion analysis and public opinion management.

Discussions

The discussion comprises four parts: (1) the convergence analysis on the opinion-trust coevolution process; (2) a quantitative effectiveness analysis with other existing methods to verify the necessity of the proposed manipulative behavior regulation mechanism and the machine moderator; (3) a qualitative comparative analysis with other existing methods to illustrate the methodological innovations of this work; (4) limitations and future work. To avoid repetitive descriptions, the parameters in this chapter are the same as section 2.3.

Table 5 The consensus-reaching speed under different parameters.

Iterations			Initial social network connectivity		Trust evolution speed: Slow (0.8, 0.1)	Medium (0.7, 0.2)	Fast (0.6, 0.3)
GDM scale	Small	(30)	Sparse	0.2	3.59	3.12	2.73
			Normal	0.5	2.29	2.31	2.27
			Intensive	0.8	1.98	1.92	2.17
	Medium	(50)	Sparse	0.2	10.26	7.98	5.39
			Normal	0.5	3.87	4.84	4.69
			Intensive	0.8	2.97	3.43	3.73
	Large	(100)	Sparse	0.2	20.07	16.69	11.55
			Normal	0.5	19.97	16.12	11.31
			Intensive	0.8	15.1	14.58	11.04

Convergence analysis of coevolution. The impacts of the three parameters of the GDM scale, the initial social network connectivity, and the trust evolution speed on opinion–trust coevolution and the CRP were explored. The consensus-reaching speed can be expressed by the number of iterations required to reach a consensus.

The parameter values were as follows: (i) the GDM scale P could be small (30), medium (50), or large (100); (ii) the initial social network connectivity could be sparse (0.2), normal (0.5), or intensive (0.8); and (iii) the social network evolution speed (r, s) could be slow $(0.8, 0.1)$, medium$(0.7, 0.2)$, or fast $(0.6, 0.3)$. Random initial opinions and a random initial social network were used to simulate coevolution 1000 times. The simulation results are reported in Table 5. The proposed opinion-social network coevolution process was convergent. And only the number of rounds to reach consensus will be different.

Effectiveness analysis. The proposed human-machine collaborative mechanism aims to improve the group consensus level with a low group manipulation tendency. To explore whether the method is effective, the following five simulation experiments were designed:

Group 1: Natural coevolution without added intervention;

Table 6 The results of the one-way ANOVA.

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F Value	*p*-value
Between groups	160.0	4	40.00	12.31	<0.001
Within groups	308.0	95	3.24		
Total	468.0	99			

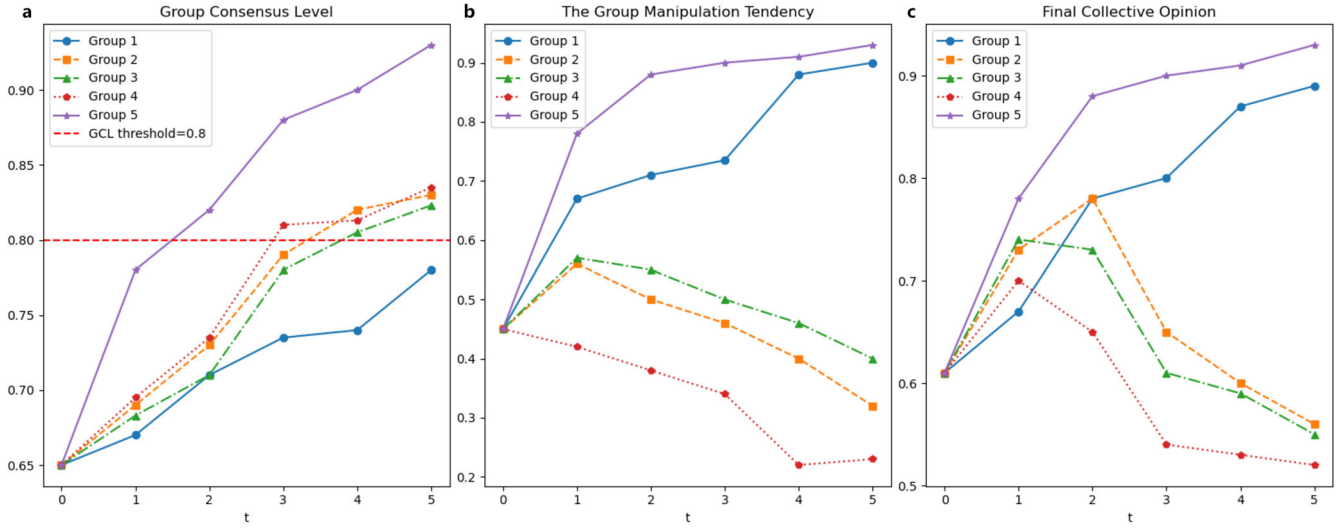


Fig. 12 Effectiveness simulation results. a Group consensus level comparison between five experiment groups; **b** Group manipulation tendency comparison between five experiment groups; **c** Collective opinion comparison between five experiment groups.

Group 2: Managing manipulative behaviors through the weight adjustment method proposed in reference (Liu et al., 2023);

Group 3: Managing manipulative behaviors through the opinion modification method proposed in reference (Gong et al., 2024);

Group 4: Managing manipulative behaviors through the trust relationship recommendation method proposed in this study;

Group 5: Reaching consensus through the RL method proposed in reference (Hassani et al., 2023).

Without loss of generality, each experimental group was adjusted to comprise 10% manipulators with a manipulation tendency degree of "1". The manipulative behavior management experiments were repeated 500 times. First, a one-way ANOVA was used to measure the significance of the effect of the behavioral management approach on the speed of consensus attainment, i.e., the number of rounds of consensus attainment, across the five experimental groups. $F = 12.31$ indicates a strong difference between group means compared to internal variance. And $p < 0.001$ shows that the differences in consensus speed across the five groups are statistically significant. The results of the one-way ANOVA are displayed in Table 6.

Furthermore, each group of simulation experiments was repeated 500 times and the experimental results were averaged. Because this research focuses on manipulation, the validity of the methodology is analyzed by comparing the group consensus level, the group manipulation tendency, and the final collective opinion, the comparative results of which are displayed in Fig. 12a, b, and c, respectively.

Compared with the control group (Group 1), the three mainstream manipulative behavior management methods (Groups 2, 3, and 4) were able to regulate manipulative behaviors to a certain extent, enhance the consensus reaching speed, and correct the group preference to around 0.5. The proposed method (Group 4) was more efficient and significantly reduced the tendency of group manipulation compared to the methods proposed in references (Liu et al, 2023) and (Gong et al., 2024) (Groups 2 and 3, respectively). This may be because opinion modification and weight adjustment can only regulate a single individual, whereas the proposed social network trust relationship recommendation can extend the influence to an individual's circle of trust, and may even affect the whole group, which has a great influence on the collective opinion. Therefore, in LGDM problems, it is more effective to recommend social network trust relationships to regulate manipulative behaviors and reduce the tendency of manipulation by the proposed method.

Regarding the method proposed in reference (Hassani et al., 2023) (Group 5), the introduction of RL algorithms caused the consensus level to approach 0.8 in the first round of discussion, which intensely sped up the CRP. However, the group manipulation tendency was as high as 0.9. and the final collective opinion converged to 0.9, which was completely influenced by the added anomalous data. This indicates that it is unreasonable to promote the CRP through machine intelligence without considering human behavior. Moreover, when abnormal data were added artificially, this human-machine collaborative model was less risk-resistant. Although a consensus could be reached quickly, the final decision-making results were unreasonable and the decision-making quality was lower. In contrast, the proposed method (Group 4) utilizes the objective and impartial characteristics of machine intelligence, which can accelerate consensus reaching while considering manipulative behaviors, ultimately providing a risk-resistant capability to ensure the reliability of the group's final collective opinion. The proposed collaborative model of human decision-making and machine supervision further strengthens the degree of human-machine collaboration, giving fair play to their respective advantages and effectively improving the quality of GDM.

Comparative analysis. To highlight both the commonalities and distinctions, the proposed approach is evaluated alongside ten

Table 7 The comparative analysis with existing methods.

Reference	Research problem	Behavior Identification	Behavior Management	Human-machine collaboration
Sun et al.	Subgroup weight manipulation	Power index	Weight punishment, feedback model	×
Wu et al.	Individual and subgroup manipulation	Clustering	Feedback model	×
Dong et al.	Trust relationship manipulation	Clustering	Leadership improvement	×
Chen et al.	Consensus manipulation	Value measurement	Adjust initial opinion	×
Liu et al.	Subgroup manipulation	Subgroup manipulation tendency	Anti-Manipulation Weight Function	×
SASAKI et al.	Strategic manipulation	Game theory	×	×
Wang et al.	Consensus achievement	×	×	Deep RL
HASSANI et al.	Weight adjustment	×	×	RL
Hou et al.	Herd behavior	Behavior consistency	Weight punishment	Clustering
Proposed method	**Manipulation**	**Manipulation tendency**	**Trust adjustment**	**RL**

representative state-of-the-art GDM methods. A detailed comparison is provided in Table 7.

(1) Compared with (Sun et al, 2023), (Wu et al, 2021), (Dong et al, 2021), (Chen et al, 2023), (Liu et al, 2023), (Sasaki et al, 2023). Scholars at home and abroad have designed different identification rules for different manipulative behaviors, such as weight manipulation, trust relationship manipulation, strategic manipulation, and consensus manipulation. The research is more detailed, but there is less literature on managing manipulative behaviors through social network trust relationships, and only reference (Dong et al, 2021) has conducted a preliminary study on the impact of manipulative behaviors through leadership relationships. This paper is innovative in abating manipulation tendency through the feedback adjustment of social network trust relationship.

(2) Compared with (Wang et al, 2022), (Hassani et al., 2023), (Hou et al, 2025). There has been some research on scholars' enhancement of GDM through machine intelligence, such as solving the problems of consensus enhancement and weight adjustment. However, there are fewer studies on managing human behavior from the perspective of human-machine collaboration, and although reference (Hou et al, 2025) studies the herd behavior, the method used is clustering, and its adaptivity is still to be improved. In this paper, training machine facilitators to manage manipulation behavior through RL, so as to realize dynamic human-machine collaboration still has certain research value.

Limitations and future work. While this study represents an initial attempt to explore theoretical and practical aspects of public opinion management, it is not without limitations. First, the proposed opinion-trust coevolution model, while effective in simulating group decision-making processes, is based on simplified assumptions regarding individual behavior and social influence, which may not fully capture the complexity of real-world public opinion dynamics. Second, the machine moderator was trained and tested within a simulated environment based on a specific scenario (i.e., the torrential rainstorm in Fujian Province), and the model involves relatively high computational complexity. These factors may limit the generalizability and practical applicability of the proposed approach to other types of emergency events or more complex social environments.

Future work can be extended in several directions. First, trust in real-world social networks often exhibits propagation characteristics, whereby individuals extend or transfer trust to others based on existing relational ties. Such propagation plays a critical role in the diffusion of influence and the evolution of group opinions. Therefore, an important direction for future research lies in integrating trust propagation mechanisms into opinion-trust coevolution model. This would enable a more comprehensive representation of opinion and trust dynamics. Then, deploying the proposed human-machine collaborative decision-making mechanism in real-world public opinion management platforms and conducting field-based validations with user feedback will help translate theoretical advances into practical implementations and continuous system improvement.

Conclusion

This work proposed a human-machine collaborative consensus reaching mechanism that considers manipulation tendency. This mechanism can greatly improve the consensus level and reliability of collective public opinion in emergency situations. The contributions of this research can be summarized in the following three points.

(1) Considering the objectivity, impartiality, and data sensitivity of machines, a new collaborative decision-making mechanism to manage manipulative behavior was proposed. Specifically, the human group discusses and assembles collective opinions, and the machine moderator learns human behavioral data and feeds back adjustment parameters to supervise their social network. This allows for a high consensus collective opinion while controlling the manipulation of tendency.

(2) Manipulators can only influence the collective opinion through two behaviors: expressing their own opinion and influencing the opinion expression of others. With this in mind, a two-dimensional behavioral manipulation tendency measurement function based on extreme opinions and expected influence is proposed, emphasizing the importance of predicted influence in manipulation identification. Further, we measure the manipulation tendency through the neutral element parameter of the non-matrix operator and set it as a gain function to train machine moderators, which achieves some effectiveness in managing manipulative behaviors.

(3) Considering that the public is unorganized and easily influenced, and that its reliability and stability are debatable, an opinion fusion method for use in emergency environments was proposed. The proposed method not only improves the consensus of collective public opinion through coevolution but also improves the reliability of collective public opinion through RL. The result not only enhances the reliability and democracy of public opinion, but also rapidly improves the public's ability to perceive the state of affairs in emergency scenarios.

However, this work is still in its infancy, leaving ample room for exploration in the future. The opinion-trust coevolution model, while effective, is based on simplified assumptions and may not fully capture real-world complexities. Additionally, the machine moderator was tested in a specific simulated scenario, which limits its generalizability. Future research should integrate trust propagation mechanisms into the model and validate the approach in real-world public opinion platforms to ensure its practical applicability and improvement.

Data availability

No datasets were generated or analyzed during the current study.

Received: 6 February 2025; Accepted: 28 July 2025;
Published online: 21 August 2025

References

Abounacer R, Rekik M, Renaud J (2014) An exact solution approach for multi-objective location–transportation problem for disaster response. Comput Oper Res 41:83–93

Ashraf S, Garg H, Kousar M(2023) An industrial disaster emergency decision-making based on China's Tianjin city port explosion under complex probabilistic hesitant fuzzy soft environment Eng Appl Artif Intell 123:106400

Bolton C, Machová V, Kováčová M, Valášková K (2018) The power of human–machine collaboration: artificial intelligence, business automation, and the smart economy. Econ, Manag, Financ Mark 13:51–56

Cai Y, Golay MW (2023) A dynamic Bayesian network-based emergency decision-making framework highlighting emergency propagations: Illustrated using the Fukushima nuclear accidents and the Covid-19 pandemic. Risk Anal 43:480–497

Chen X, Liang H, Zhang Y, Wu Y(2023) Consensus manipulation in social network group decision making with value-based opinion evolution Inform Sci 647:119441

Dong Y, Zha Q, Zhang H, Herrera F (2021) Consensus reaching and strategic manipulation in group decision making with trust relationships. IEEE Trans Syst, man Cybern Syst 51:6304–6318

Duan J, Shi D, Diao R, Li H, Wang Z, Zhang B, Bian D, Yi Z (2020) Deep-reinforcement-learning-based autonomous voltage control for power grid operations. IEEE T Power Syst 35:814–817

Fodor JC, Yager RR, Rybalov A (1997) Structure of Uninorms. Int J Uncertain Fuzziness Knowl-Based Syst 05:411–427

Fu S, Xiao Y, Zhou H (2022) Contingency response decision of network public opinion emergencies based on intuitionistic fuzzy entropy and preference information of decision makers. Sci Rep-UK 12:3246

Gao S, Zhang Y, Liu W (2021) How does risk-information communication affect the rebound of online public opinion of public emergencies in China? Int J Environ Res Public Health 18:7760

Garai T, Garg H, Biswas G (2023) A fraction ranking-based multi-criteria decision-making method for water resource management under bipolar neutrosophic fuzzy environment. Artif Intell Rev 56:14865–14906

Gong G, Zhou X, Zha Q (2024) Managing fairness and consensus based on individual consciousness of preventing manipulation. Inform Fusion 102:102047

Haesevoets T, De Cremer D, Dierckx K, Van Hiel A (2021) Human-machine collaboration in managerial decision making. Comput Hum Behav 119:106730

Han R (2023) More talk, more support? The effects of social network interaction and social network evaluation on social support via social media. Psychol Res Behav 16:3857–3866

Hassani H, Razavi-Far R, Saif M, Herrera-Viedma E (2023) Reinforcement learning-based feedback and weight-adjustment mechanisms for consensus reaching in group decision making. IEEE Trans Syst Man, Cybern Syst 53:2456–2468

He Q, Lv Y, Wang X, Li J, Huang M, Ma L, Cai Y (2023) Reinforcement-learning-based dynamic opinion maximization framework in signed social networks. IEEE Trans Cogn Dev Syst 15:54–64

Hengsheng Z, Rui Z, Quantao W, Haobin S, Hwang K (2022) An adaptive algorithm for consensus improving in group decision making based on reinforcement learning. J Chin Inst Eng 45:161–174

Hong W, Gu Y, Wu L, Pu X (2023) Impact of online public opinion regarding the Japanese nuclear wastewater incident on stock market based on the SOR model. Math Biosci Eng 20:9305-9326

Hou Y, Xu X, Pan B (2025) Herd behavior identification based on coevolution in human–machine collaborative multi-stage large group decision-making. Inform Sciences 689:121511

Hui F (2022) Research on the construction of emergency network public opinion emotional dictionary based on emotional feature extraction algorithm. Front Psychol 13:857769

Jia P, MirTabatabaei A, Friedkin NE, Bullo F (2015) Opinion dynamics and the evolution of social power in influence networks. SIAM Rev 57:367–397

Lettieri N, Guarino A, Zaccagnino R, Malandrino D (2023) Keeping judges in the loop: a human–machine collaboration strategy against the blind spots of AI in criminal justice. Soft Comput 27:11275–11293

Li P, Xu Z, Zhang Z, Li Z, Wei C (2023) Consensus reaching in multi-criteria social network group decision making: A stochastic multicriteria acceptability analysis-based method. Inform Fusion 97:101825

Li W, Guo C, Deng Z, Liu F, Wang J, Guo R, Wang C, Jin Q (2023) Coevolution modeling of group behavior and opinion based on public opinion perception. Knowl-based Syst 270:110547

Lillywhite B, Wolbring G (2022) Emergency and Disaster Management, Preparedness, and Planning (EDMPP) and the 'Social': A scoping review. Sustainability 14:13519

Liu B, Zhou Q, Ding R, Palomares I, Herrera F (2019) Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. Eur J Oper Res 275:737–754

Liu P, Dong X, Wang P, Du R (2024) Managing manipulation behavior in hydrogen refueling station planning by a large group decision making method with hesitant fuzzy linguistic information. Inform Sci 652:119741

Liu Y, Wei G, Liu H, Xu L (2022) Group decision making for internet public opinion emergency based upon linguistic intuitionistic fuzzy information. Int J Mach Learn Cyb 13:579–594

Maruyama S, Moriya S (2021) Newton's Law of Cooling: Follow up and exploration. Int J Heat Mass Transfer 164:120544

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015) Human-level control through deep reinforcement learning. Nature 518:529–533

Morgan TKKB, Fa AuiTN (2018) Empowering indigenous voices in disaster response: Applying the Mauri Model to New Zealand's worst environmental maritime disaster. Eur J Oper Res 268:984–995

Mostafa MM (2021) Information diffusion in halal food social media: a social network approach. J Int Consum Mark 33:471–491

Ni W, Shen Q, Liu T, Zeng Q, Xu L (2023) Generating textual emergency plans for unconventional emergencies — A natural language processing approach. Safety Sci 160:106047

Ni W, Shu J, Song M (2018) Location and emergency inventory pre-positioning for disaster response operations: Min-Max Robust model and a case study of Yushu earthquake. Prod Oper Manag 27:160–183

Rathi R, Khanduja D, Sharma SK (2017) A fuzzy-MADM based approach for prioritising Six Sigma projects in the Indian auto sector. Int J Manag Sci Eng 12:133–140

Ren S, Gong C, Zhang C, Li C (2023) Public opinion communication mechanism of public health emergencies in Weibo: take the COVID-19 epidemic as an example. Front Public Health 11:1276083

Sasaki Y (2023) Strategic manipulation in group decisions with pairwise comparisons: A game theoretical perspective. Eur J Oper Res 304:1133–1139

Scuotto V, Del Giudice M, Peruta MRD, Tarba S (2017) The performance implications of leveraging internal innovation through social media networks: An empirical verification of the smart fashion industry. Technol Forecast Soc Change 120:184–194

Sherif M, Hovland CI (1961) *Social judgment: Assimilation and contrast effects in communication and attitude change*. Yale Univer. Press: Oxford, England

Song L, Gong M, Wu C (2013) Realization of urban emergency decision-making aid system. Sci Surv Mapp 38:190–202

Sun Q, Wu J, Chiclana F, Wang S, Herrera-Viedma E, Yager RR (2023) An approach to prevent weight manipulation by minimum adjustment and maximum entropy method in social network group decision making. Artif Intell Rev 56:7315–7346

Wang M, Liang D, Xu Z (2022) Consensus achievement strategy of opinion dynamics based on deep reinforcement learning with time constraint. J Oper Res Soc 73:2741–2755

Wilson HJ, Daugherty PR (2018) Collaborative intelligence: Humans and AI are joining forces. Harvard Business Review 96(4):115–123

Wu J, Cao M, Chiclana F, Dong Y, Herrera-Viedma E (2021) An optimal feedback model to prevent manipulation behavior in consensus under social network group decision making. IEEE Trans Fuzzy Syst 29:1750–1763

Wu J, Chiclana F, Herrera-Viedma E (2015) Trust based consensus model for social network in an incomplete linguistic information context. Appl Soft Comput 35:827–839

Xiong K, Dong Y, Zha Q (2024) Managing strategic manipulation behaviors based on historical data of preferences and trust relationships in large-scale group decision-making. IEEE Trans Fuzzy Syst 32:1479–1493
Xiong W, Wang C, Ma L (2023) Partner or subordinate? Sequential risky decision-making behaviors under human-machine collaboration contexts. Comput Hum Behav 139:107556
Xue Y, Qiyu X, Juan Y, Yujia Y, Kang T, Hao C (2024) Research on Influencing Factors and Mechanisms of Human-Machine Safety Collaboration Behavior in Coal Mines Based on DEMATEL-ISM. SAGE OPEN 14
Yager RR, Rybalov A (1996) Uninorm aggregation operators. Fuzzy Set Syst 80:111–120
Yang W, Zhang L, Shi J, Lin R (2024) New consensus reaching process with minimum adjustment and feedback mechanism for large-scale group decision making problems under social trust networks. Eng Appl Artif Intel 133
Zhao J, He H, Zhao X, Lin J (2022) Modeling and simulation of microblog-based public health emergency-associated public opinion communication. Inf Process Manag 2:102846
Zhao Y, Xu M, Dong Y, Peng Y (2021) Fuzzy inference based Hegselmann–Krause opinion dynamics for group decision-making under ambiguity. Inform Process Manag 58:102671

Acknowledgements

This work was supported by the major project of National Natural Science Foundation of China Major Project (72293574, 72091515), Key R&D Projects of Hunan Province(2024AQ2016) and the Major Program of Xiangjiang Laboratory (No.23XJ01005).

Author contributions

The author's contributions to this work are presented as follows.1.Yuzhou Hou: Conceptualization, Methodology, Software, Data curation, Investigation, Formal analysis, Writing original Draft.2.Xuanhua Xu: Review & Editing, Validation, Supervision, Funding acquisition.3.Zongrun Wang: Data curation, Revision, Investigation.4.Weiwei Zhang: Data curation, Revision, Investigation.

Competing interests

The authors declare no competing interests.

Ethical approval

This study did not involve any experiments with human participants or animals performed by any of the authors. This study is based entirely on social media public opinion data. Therefore, ethical approval was not required. The study complied with the ethical standards outlined in the Declaration of Helsinki and its later amendments.

Informed consent

This study did not involve human participants, and thus, no informed consent was required.

Additional information

Correspondence and requests for materials should be addressed to Xuanhua Xu.