International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 486-503

Article

Deep Behavioral Recognition Model Construction for Live Streaming Reward Fraud - Fusing Music Sentiment Analysis and User Interaction Timing Graph Neural Networks

Yihan Sun and Hyuntai Kim *

Department of Music, School of Arts and Physical Fitness, Sejong University, Seoul 05006, Republic of Korea; 18098888398@163.com

Abstract: The rapid development of the network live broadcasting industry has given rise to the economic model of reward, but it also brings the problem of frequent fraud. Anchor through false identity packaging, acting on behalf of playing, software beauty and other means to mislead users to irrational reward, the platform lacks an effective regulatory mechanism, the user's rights and interests are seriously damaged. Traditional fraud detection methods are difficult to effectively identify complex user interaction behavior patterns and emotional manipulation, and there is an urgent need to build an intelligent detection model that integrates multi-dimensional information to protect consumer rights and interests. In view of the problem of identifying fraudulent behavior of network live streaming bounty, the research constructed a graph neural network deep recognition model that integrates music sentiment analysis and user interaction time sequence. The model extracts audio features of the live broadcast through the music sentiment analysis module, and uses the Mel frequency cepstrum coefficient for sentiment classification; constructs a multi-view graph neural network architecture, uses the cosine similarity and K-nearest neighbor algorithm to construct node relationships, and fuses the embedded vector information from different viewpoints through the attention mechanism; and designs a five-classification detection framework to identify the behavioral types of normal interactions, irrational consumption frauds, shoddy product frauds, false information frauds, and misconception frauds. fraud and misperception fraud and other behavior types. Experimental results on Yelp and Amazon datasets show that the model improves 0.1298, 0.2186, and 0.2364 in F1 metrics compared to GCN, GAT, and GraphSAGE, respectively, and achieves 98.95% recall and 99.14% precision in the five-classification detection of webcaster fraud. The model effectively solves the problems of sample imbalance and complex interaction relationship identification in live streaming reward fraud detection, and provides technical support for risk control of webcasting platforms.

Keywords: live streaming reward fraud; graph neural network; music sentiment analysis; user interaction; attention mechanism; deep recognition model

1. Introduction

With the rapid development of the Internet live broadcasting industry, the live broadcasting room has become an important platform for people to socialize and entertain, obtain information, and shop and consume [1-2]. However, at the same time, the phenomenon of live streaming bounty fraud has become increasingly rampant, bringing great economic losses and mental harm to the majority of users [3-4].

There are many ways of live streaming reward fraud, mainly including the following: (1) emotional inducement scam, the fraudster will disguise himself as a gentle and considerate, understanding image on the live streaming platform, establish emotional contact with the victim, and obtain the victim's trust.

Then the victim is induced to make rewards for various reasons, and once the victim is caught in the emotional trap, it is easy to make a large number of rewards without realizing it [5-8]. (2) False identity fraud, some unscrupulous elements will pose as well-known anchors, stars or rich second-generation, etc., and release false information in the live broadcast to attract the attention of the audience, they claim that they have an internal channel to obtain high returns, or precious investment opportunities, to entice the audience to participate and make rewards [9-12]. (3) Prize winning scam, the scammers set up a fake prize-winning link in the live broadcast, claiming that viewers have a chance to win the prize as long as they participate in rewarding, however, when viewers reward, they can't get the so-called prizes at all [13-15]. (4) Induced to brush gifts to upgrade scam, some anchors will induce viewers to brush gifts to upgrade the level, claiming that the higher the level can get more privileges or better services, some viewers in order to satisfy the vanity or get the anchor's special attention, they will keep brushing the gifts to upgrade, but in fact, these anchors colluded with fraudsters to use the audience's reward for personal gain [16-19]. Therefore, in order to effectively combat live streaming reward fraud and protect users' rights and interests, it is particularly important to model and analyze live streaming fraud data [20-21]. By constructing a deep behavioral recognition model, we can gain a deeper understanding of the laws, characteristics and trends of fraudulent behavior, thus providing strong support for the development of accurate prevention strategies and combat measures [22-23].

The booming development of the network live broadcasting industry has reshaped the ecological pattern of digital entertainment, and live broadcasting reward as the core business model has created huge economic value, but the accompanying fraud problem is becoming increasingly prominent. There are multi-level fraud risks in the current live broadcasting ecology: the anchor level induces users to consume irrationally by means of identity forgery, skill counterfeiting, emotional manipulation, etc.; the platform level has an imperfect regulatory mechanism and a limited ability to identify fraudulent behaviors; and the user level lacks an effective protection mechanism, which is easily guided by emotions to produce impulsive consumption. Existing fraud detection technology mainly relies on rule matching and shallow machine learning methods, it is difficult to capture the complex multimodal information and dynamic interactive relationships in the live scene. Traditional methods have limitations in dealing with the fusion of multi-dimensional information such as audio emotional features, user behavior temporality, and social network relationships, resulting in insufficient detection accuracy and generalization ability. Graph neural network technology provides a new idea to solve this problem, and its powerful relational modeling ability can effectively deal with the complex relational structure in the user interaction network, but how to organically fuse the audio emotion information with the graph structure information is still a technical difficulty.

Based on the above analysis, this study proposes a deep behavior recognition model that integrates music sentiment analysis and user interaction timing graph neural network. Firstly, the music sentiment analysis module is constructed, and the audio features of the live broadcasting room are extracted and sentiment classified through the Mel frequency cepstrum coefficient; then the multi-view graph neural network architecture is designed, and the dual-view graph structure is constructed by using the original interaction relationship and the feature similarity; finally, the multi-modal information is fused through the attention mechanism, and the end-to-end fraud detection framework is constructed. This research provides technical support for the risk control of live broadcasting platforms and the protection of user rights and interests.

2. Live Streaming Bounties and Live Streaming Bounty Fraud Determinations

2.1. Rewards for Live Broadcasts

Live streaming bounty refers to the behavior of users in the process of network live broadcasting to the anchor to send gifts and other ways to reward. Users to buy gifts in virtual currency, after the reward user identity from the performance of the viewer into a consumer, the anchor of the reward behavior to be a positive response.

2.1.1. Types of Live Broadcast Bounties

There are five main types of live streaming bounties:

One is bounty as encouragement. Performers produce high-quality content, perform brilliantly or have desirable features such as dry goods, users will reward them as encouragement.

The second is rewarding with a sense of accomplishment. Users believe that the reward will be a source of income for the anchor, and have a sense of achievement or helpfulness for the performer's reward. Fans pay for the anchor they support and have a positive experience of accomplishment after rewarding.

Third, show-off bounty. Reward is to show others their strength or ability, there is a kind of bragging mentality, and will get the title or ranking, there is a privileged feeling of looking down on others, specific power to bring a good self-experience.

The fourth is pure fun or interest in rewarding. There is no specific purpose, users reward casually or follow the trend of reward, reward as fun.

Fifth, purposeful rewarding. Users get the contact information of the anchor through rewarding. Brush gift on demand talent for interaction.

2.1.2. Characteristics of Live Broadcast Bounties

One is non-compulsory. No matter what kind of psychological origin of reward behavior, behind what kind of influence mechanism, are carried out according to the individual will, is the individual dynamic behavior.

Second, interactivity. Live reward is the process of interaction between the user and the anchor in the live broadcast.

Third, entertainment. Users reward behavior from the fun, is to pass the time of daily life entertainment behavior, the anchor performance brush different amount of gifts to encourage, with pan-entertainment characteristics, and then form entertainment culture.

Fourth, virtualized network socialization. Network live broadcasting platform to build a social network, live bounty constructed a diversified virtualized social interaction mode.

Fifth, the fan economy. Live streaming reward is the embodiment of the platform economy, is built on the basis of user attention or fan support for consumer behavior, is based on the new economic model of the digital economy era of emotion.

Network live broadcasting platform will live reward as a business model, through various ways to stimulate consumers, create an atmosphere to stimulate users to reward, is its cultivation of consumer behavior payment mode.

For the anchor, getting reward is getting revenue. Become a platform anchor threshold is low, there are cell phones, microphones and other live equipment, there are rooms or outdoor and other suitable live space environment can be live, with easy access. Generally need to set the image for packaging, this persona or real or virtual, live will maintain the set image, is a kind of emotional labor.

2.2. Practical Logic of Webcast Rewarding

(1) The product of the development of network technology in the information age

In the contemporary era where emotional deficiencies are becoming more and more common and individuals pursuing personalization are beginning to spontaneously seek new communities, the birth of webcasting undoubtedly fits the needs of contemporary young people pursuing the reconstruction of emotional order.

In recent years, WeChat, Alipay and other payment methods have gradually overtaken cash payments to occupy the mainstream position, and paper money has gradually faded from people's view. Behind this phenomenon is a strong digital economy information technology in support, information technology and the gradual maturation of the digital economy has become the objective technical conditions for people to dare to carry out network live reward.

(2) The broadening of emotional expression and social ways

Different from the emotional interaction based on real social interaction in the past, webcasting and its reward present new emotional construction characteristics. Network live broadcasting has broadened the scope of socialization, and people can intersect with people they have never met through live broadcasting and rewarding, and the existence of distance makes it easier for people to express their emotional needs. The network live broadcast reward makes the host's live broadcast behavior become a kind of “emotional labor”, the audience's reward is also with the nature of “emotional consumption”. Under the mutual collision and in-depth interaction between the two, the process of mutual emotional expression and demand fulfillment has undergone profound changes. Webcasting is characterized by authenticity, co-timeliness, commerciality and interactivity. Among them, interactivity is directly related to the establishment of individual emotional order.

2.3. Theoretical Logic of Webcast Rewarding

(1) Anchor's Perspective

In contemporary times, one of the reasons for the growing popularity of webcast bounties lies in the fact that virtual currencies expand the distance between the bounty giver and the recipient. In the process of service, the emotions conveyed to others depend on human emotional labor, i.e., “inducing or suppressing one's own feelings in order to maintain an appropriate expression in order to produce an appropriate mental state in others.” Thus, in the sociological theory of emotion, the anchor's live broadcasting behavior takes emotion as one of the contents of serving others, which belongs to a kind of emotional labor.

(2) Audience perspective

Another reason for the growing popularity of network live broadcasting is the increasing richness of the live broadcast form and live content. For the live content, the performance of life chores as small as eating and sleeping, to the immersive live broadcast of extreme sports, almost covering all the life experiences that people can think of.

(3) Reconstruction of emotional order under the interaction between viewers and anchors

In order to meet the audience's emotional needs and gain attention, reward and other rewards, anchors often adopt a series of "emotional operation" strategy to construct the value of the action of emotional rationalization mechanism. From the anchor's commodity (i.e., emotion) production to the audience's consumption of this process is actually the anchor constantly obtain self-identity process.

The anchor must communicate with the audience and produce certain emotional responses to the messages and behaviors of others, such as gratitude, expectation, and dissatisfaction. Maintaining the emotional connection between the anchor and the audience is an important factor for the survival of live broadcasting, and rewarding gifts is an effective way to maintain the emotional connection.

If the anchor's behavior is understood as "emotional labor", then the audience's reward accordingly becomes a kind of "emotional consumption". Emotional consumption is "a consumer behavior in which an individual obtains emotional commodities through market-oriented exchange, so that he or she can obtain emotional satisfaction and emotional support".

2.4. Direct Broadcast Reward Fraud

There are four elements of civil fraud: subjective intent, fraudulent conduct, and dual causation. It should also be noted that the criteria for the determination of consumer fraud are different from those for the determination of civil fraud in doctrine and practice.

2.4.1. Determination of Fraud by Webcast Hosts

(1) The subjective intent of the anchor fraud

Regardless of the type of anchor, "distinctive features" is the basis for its promotion, but also many users are willing to become its fans, the key to their gifts.

Anchor to commit fraud, usually directly related to the "distinctive characteristics" of this key factor. Therefore, it can be said that the host who committed fraud, at least with the hope that users fall into the misperception (the host has the characteristics of differentiation), and hope that the users based on this misperception of the indirect intent to give their virtual gifts. So far, the implementation of fraudulent behavior of the anchor clearly has the subjective double intent of fraud.

(2) Anchor using false means of live constitutes fraudulent behavior

Specific performance is as follows:

First, the use of plug-ins, loopholes for live.

Secondly, on behalf of playing, acting on behalf of alternative behavior.

Third, the use of software high strength to change their own physical characteristics for live broadcast.

(3) Anchor fraud is the reason for the user's misperception

For the anchor, especially the new anchor, if you want to get the number of fans, you must need a special title, so that "passing" users have the desire to point into their own live room. When the user is attracted by the title or the picture of the live room and click into the live room, if the anchor is true to its name, the user will often stay in the live room. If the anchor's live broadcast does not exist with the "distinctive features", most users will scoff, and then turn off the live room.

(4) The results of the user's erroneous reward due to the fraud by the anchor

Users who rewarded out of "appreciation for the anchor" will fall into the misconception that the anchor has the "distinctive features" claimed by the anchor due to the anchor's fraud, and based on this misconception, they will give virtual gifts to the anchor.

2.4.2. Determination of Fraud on Webcasting Platforms

This paper emphasizes that the network live broadcasting platform bears too little risk in the process of rewarding activities. There is an obvious imbalance of rights and obligations between the webcasting platform and the user in the process of signing the contract for the sale of virtual tokens and the contract for the sale of virtual gifts paid for by virtual tokens (hereinafter referred to as the dual sale contract). There is an urgent need to increase the obligation commitment of the webcasting platform to achieve a balance of rights and obligations.

Moreover, it is obviously unreliable to rely solely on the network live broadcasting platform to deal with the offending anchor through its own intrinsic and spontaneous manner. After all, the network live broadcast platform and anchor this “athlete” in the economic interests of this is a hill of raccoons, network live broadcast platform this “referee” how can easily penalize their own people violate it. Therefore, the network live platform into the user error reward results in the responsibility of the main body is necessary.

(1) The network live platform operator.

(2) The status of network live broadcasting platform operator is recognized.

Network live broadcasting platform of the main identity of the problem will directly affect its rights and obligations in multiple contracts, will also affect the network live broadcasting platform in the double sale contract rights and obligations, more will deeply affect the network live broadcasting platform and the relationship between the user.

(3) The network live broadcasting platform has the obligation to inform the user of the real live broadcasting situation of the anchor

(4) The network live broadcasting platform fails to fulfill the obligation to inform constitutes fraud

3. Anchor Fraud Detection Method Based on Graph Neural Network

3.1. Technical Basis

3.1.1. Music Sentiment Analysis

Firstly, the music data of each song is preprocessed, and the frequency domain features and Mel's inverted spectral features of the song are used to design the classifiers to test and evaluate the classification effect, respectively. After comparing the results of the classifiers constructed based on different audio features, the Mel frequency cepstrum coefficient (MFCC), a feature with higher accuracy, is selected as the basis for final classification to complete music emotion classification [24].

(1) Music data preprocessing. This is mainly for the popular songs in the webcast room, which are included and screened to form the webcast song list. SoX audio processing software is used to convert all MP3 songs into mono, 22000Hz WAV files.

(2) Manual calibration of music emotion classification. Since the music emotion classification is subject to subjectivity, multiple people are used to calibrate the same music clip, and the one with the largest number of people is used to complete the emotion classification of the songs.

(3) Divide the training set and test set. The preprocessed WAV files are divided into two subsets according to 8:2, the former is used as the training set to construct the classifier, and the latter is used as the test set to evaluate the performance of the classifier.

(4) Music feature extraction. In this paper, we mainly extract music features from the frequency domain, and extract the spectral features and Mel's inverted spectral features of the songs respectively.

(5) Training the classifier. Considering that this project is a multi-classification problem, a logistic regression classifier is used for music classification.

(6) Test and evaluate the classification results.

In this paper, Fast Fourier Transform is used to calculate the spectral features of each song. The Meier frequency cepstrum coefficient is a cepstrum parameter extracted from the scalar frequency domain, which describes the nonlinear characteristics of the audio signal in the frequency domain. The original spectrum consists of two parts: the envelope and the details of the spectrum. The Meier frequency cepstrum coefficients can be extracted if a set of bandpass filters are set up sequentially from low to high frequencies according to the size of the critical bandwidth, the input signal is filtered, and then the signal energy at the output of each bandpass filter is used as the basic feature of the signal. Since the feature does not depend on the nature of the signal, it has good recognition and noise immunity, and is more in line with the auditory characteristics of the human ear.

The basic process of MF-CC parameter extraction is as follows:

Step1: The continuous music signal $s(n)$ is pre-emphasized, sub-framed and windowed to obtain a frame of music signal $x(n)$.

Step2: The $x(n)$ signal is subjected to Fast Fourier Transform (FFT) to obtain the frequency domain signal, i.e., the linear spectrum $X(k)$, i.e.,:

$$X(k)=\sum_{n=0}^{N-1}x(n)W^{nk},k=0,1,2,\cdots,N-1 \quad (1)$$

where $x(n)$ is the input speech signal and N is the number of points of the Fourier transform.

Step3: The linear spectrum $X(k)$ $H_m(k),0\le m\le M$, M

is the number of filter banks. The center frequency of each filter is $f(m)$. The center frequencies of the triangular filter banks are equally spaced in the Meier frequency domain. From this, the center frequency distance $\Delta Mel = \frac{f_{\max}}{k+1}$ between two adjacent delta bandpass filters can be calculated. Each bandpass filter bank passes the parameters as:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{2(k - f(m-1))}{(f(m+1) - f(m-1))(f(m) - f(m-1))}, & f(m-1) \le k \le f(m) \\ \dfrac{2(f(m+1) - k)}{(f(m+1) - f(m-1))(f(m) - f(m-1))}, & f(m) \le k \le f(m+1) \\ 0, & k \ge f(m+1) \end{cases} \tag{2}$$

Step4: The energy spectrum $s(m)$ on Mel coordinates is output via the filter bank as:

$$s(m) = \ln\left(\sum_{K=0}^{N-1} | X_a(k) | H_m(k)^2\right), 0 \le m \le M \tag{3}$$

Step5: The Mel frequency cepstrum $C(n)$ in the frequency domain is obtained by discrete cosine inverse transform:

$$C(n) = \sum_{K=0}^{N-1} s(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), n = 12, \cdots, L \tag{4}$$

where L refers to the order of MFCC coefficients. Since the MFCC coefficients in the first and last dimensions have a greater discriminative performance for speech, the first 12-dimensional MFCC coefficients and the first 12-dimensional first-order difference coefficients are used as the feature parameters in this paper.

3.1.2. Graph Neural Networks

Graph neural network is a popular research direction in the field of deep learning in recent years, which is an extension of the traditional deep learning model and is specifically designed to process graph-structured data. Graph-structured data exists in a wide range of real-life domains, such as social networks, molecular structures, and transportation networks. In these scenarios, nodes and edges represent entities and relationships between entities, respectively, and graph neural networks are able to effectively capture the complex patterns in such relational data [25-26].

The algorithmic principles of graph neural networks are mainly based on graph theory and deep learning. Its core idea is to transform the nodes and edges in the graph structure data into computational units in the neural network, and capture the complex relationships in the graph structure by iteratively updating the embedded representations of the nodes. The basic operation of a graph neural network consists of two stages: message passing and graph convolution. This process demonstrates how the GNN can effectively utilize the structural information of the graph.

(1) Message passing: used to update the embedded representation of nodes. In each iteration step, each node receives messages from its neighboring nodes and updates its own state based on these messages. The message passing process can be represented as:

$$h_v^{t+1} = f\left(h_v^t, \sum_{u \in N(v)} m(h_v^t, h_v^t, e_{vu})\right) \tag{5}$$

(2) Graph Convolution: draws on the convolution operation in convolutional neural networks. The goal of graph convolution is to update the embedding representation of a node by aggregating information from its neighbors. A common graph convolution operation can be represented as:

$$h_v^{t+1} = \sigma \geq \left(\sum_{u \in N(v)} W h_u^t \right) \quad (6)$$

where W is the learnable weight matrix and $\sigma(\cdot)$ is the activation function. This operation can be understood as a weighted summation of the embedded representations of node v and its neighboring nodes, and a nonlinear transformation by the activation function. In practice, stacking graph convolutional layers is an effective strategy to help improve the overall performance of graph neural networks.

By iteratively performing message passing and graph convolution operations, graph neural networks can progressively update the embedding representations of nodes to capture complex relationships in graph structures. Ultimately, these embedding representations can be used for a variety of graph-level tasks, such as node classification, graph classification, link prediction, etc. In summary, graph neural networks are able to capture complex relationships in graph structures through operations such as message passing graph convolution.

3.1.3. Interactive Information Data Preprocessing

(1) Interaction information cleaning and preprocessing

The graph neural network-based approach for webcasting (social media) interactive text sentiment analysis combines text cleaning and preprocessing steps, aiming to deal with heterogeneous and noisy text data on the live streaming platform, so as to improve the accuracy and reliability of sentiment analysis. Text cleaning is one of the necessary steps. Since live media texts usually contain a large amount of noise, emoticons, abbreviations, spelling errors, and other non-normalized content, cleaning these noises can help purify the text and extract meaningful information.

Common text cleaning methods include removing special characters, processing URL links, and eliminating duplicate content. Through these cleaning steps, the interference of data noise on sentiment analysis can be reduced and the performance of the model can be improved. Meanwhile, text preprocessing is crucial for building graph structure and feature extraction. When constructing the graph structure, the social media texts can be regarded as nodes and the edges of the graph can be constructed based on the relationships between users (e.g., following, liking, commenting, etc.). This can take the user interaction behavior of the live streaming platform into consideration and capture the contextual and associative relationships between the texts.

(2) Sentiment labeling and dataset construction

Sentiment labeling is one of the key steps. Since live media texts usually do not have clear sentiment labels, it is necessary to manually label a portion of the text for sentiment classification. The annotator can classify the text into positive, negative or neutral sentiment categories based on the emotional tendencies expressed in the text. Such sentiment labeling can be based on subjective judgment or follow existing sentiment labeling specifications or label sets.

In addition, dataset construction is the process of building a training dataset based on sentiment labeling. One can choose to collect large-scale text data from social media platforms and associate it with sentiment labels. This can be achieved by utilizing social media API interfaces or crawling techniques. Care should be taken while constructing the dataset to select representative and diverse text samples to cover different topics, sentiment tendencies and text styles. In the process of dataset construction, it is also necessary to consider the balance and quality of the data to ensure that the number of samples from different sentiment categories is relatively balanced to avoid the classifier's preference for certain sentiment categories. At the same time, data quality control is needed. By combining sentiment labeling and dataset construction, a training dataset with music sentiment labels and interaction information can be obtained for training and evaluating graph neural network models.

3.2. Motivation Analysis of User Information Interaction Behavior in Webcasting

In the information ecosystem, the subject element informant is the most important component. Therefore, this paper provides an in-depth analysis of the motives of information interaction behavior between informants, i.e., anchors, users, live platforms and e-commerce enterprises, in live webcasting (live streaming bounty).

The motivation of information interaction behavior in live streaming banding is divided into the motivation of information interaction behavior between the anchor and the user, the motivation of information interaction behavior between the user and the user, the motivation of information interaction behavior between the enterprise and the platform, and the motivation of information interaction behavior between the platform and the user.

The model of information interaction behavior motivation in webcasting is shown in Figure 1.

(1) Information Interaction Behavior Motivation between Anchor and User

Social demand refers to the intrinsic need for interaction between people, which is manifested in communication between people, reciprocity, social prestige and so on.

Self-actualization refers to the state of constantly exerting one's potential through personal efforts, accomplishing set goals or ideals, and gradually tending to perfection. The anchor acts as an opinion leader in the live broadcasting room and has a certain power of speech, which is higher than his or her role status in real life, thus leading to an increase in self-fulfillment.

According to the use and satisfaction theory, it is found that entertainment and leisure motives have a significant positive influence on users' electronic consumption behavior. Anchors will bring relaxation and pleasure to some users when they show their talent, teach knowledge, and explain products in the live broadcasting room.

(2) The information interaction between users and users in the live broadcasting room helps to satisfy their social and respect needs motivation, and this communication is reflected in the form of sending comments and pop-ups between users.

(3) The common motives of information interaction between enterprises and platforms include information sharing among users and cultivating the core competitiveness of enterprises. Users' behavioral motives for platform information interaction are mainly social demand motives and preference motives.

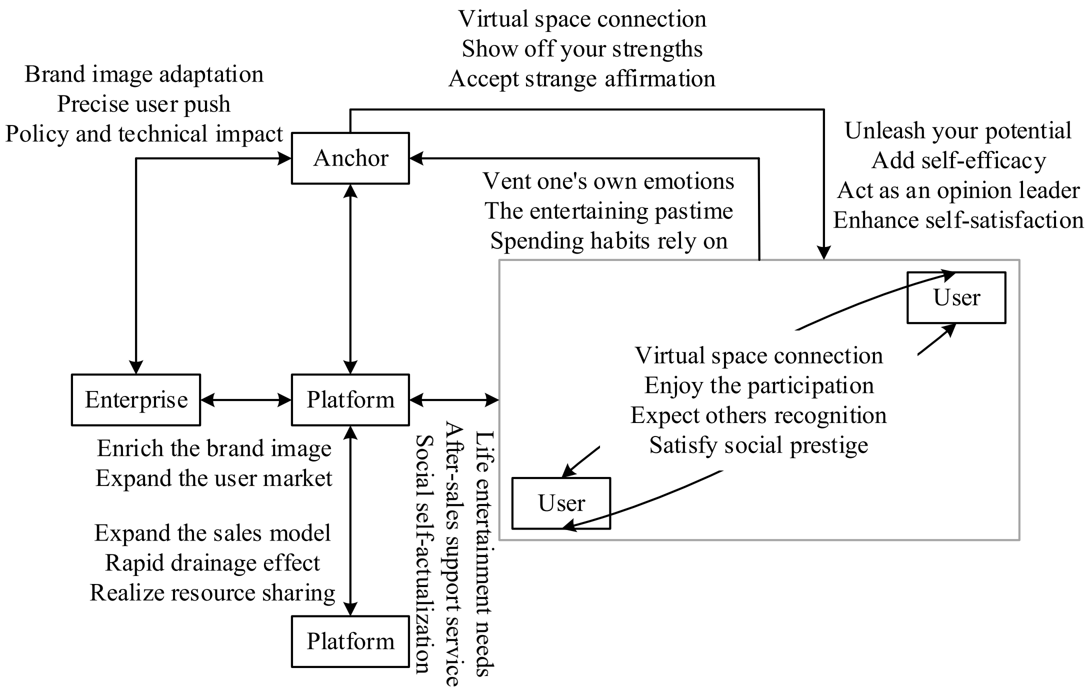

Figure 1. Network broadcast citic interest interaction behavior motivation model.

3.3. Analysis of Anchor Discourse Expression in Live Broadcasting

(1) Expression of emotions, interests, and social discourse

In the construction of network live broadcasting discourse scene, anchors often use exaggerated language and seductive discourse style to achieve the purpose of generating sensory stimulation, stimulating the desire to buy, establishing identity and maintaining intimate relationships, and obtaining good discourse communication effects.

The purpose of the speaker's emotional discourse is to maximize the sensory stimulation of online fans. Some network anchors will often be "acquaintances" identity discourse expression, with affectionate greetings, greetings and other ways to establish emotional ties with the audience, to bring themselves closer to the audience's psychological distance and emotional distance, prompting the audience to put their personal emotions in the consumption process, so as to realize the production of sensory stimulation, increase the emotional experience of the discourse expression. Language purpose.

From the perspective of discourse domain theory, network anchor discourse expression is centered on a specific discourse purpose, trying to maximize the audience's desire to buy through the construction of interest discourse expression, and then realize the sale of goods and benefit from it.
In the network live broadcast, the network anchor will use special discourse expression to establish a close relationship with the audience, to narrow the identity and psychological distance between himself and the audience, so as to improve the audience's recognition and intimacy. For example, webcasters will express themselves in the way of chatting among friends, and increase the intimacy and familiarity of both sides by using familiar language such as “thank you, baby fans” and “my friends”.

(2) Induced consumption behavior

Live streaming reward is a kind of activity that comes along with the development of self-media, and it is also a kind of social behavior of “social+experience+consumption”, which brings a special social consumption phenomenon. From the perspective of language domain theory, in order to better realize the group psychological identity of the language purpose, the anchor often adopts the “use and satisfaction” as the core purpose of the language, to guide the audience's certain consumption “expectations”. As an official media said, network anchors use exaggerated expression, adopt the language style for the purpose of maintaining intimate relationship and identity construction, carry out alternative publicity, and achieve the effect of induced consumption.

From the perspective of discourse field theory, in the live broadcast discourse field, consumers are no longer simple viewers, but become participants in live broadcast activities. This implicit social role shift of consumers has a greater impact on live broadcast activities. Driven by the psychology of blind obedience, consumers are prone to produce the consumer impulse of “everyone gives away” and implement the induced consumption behavior. In the case of social role replacement, consumers are prone to produce network live broadcasting “protagonist” psychology, unthinkingly make reward behavior, impulsive consumption.

3.4. Anchor Fraud Detection Algorithm Incorporating Multiple Factors

Fraud detection faces problems such as sample imbalance and intricate relationships between fraudulent nodes, in order to cope with this challenge, this paper proposes a fraud detection algorithm based on multi-view graph neural network.

(1) Clustering-based viewpoint information construction algorithm

For the graph dataset $G=(V,E)$, V is the node dataset of the graph G and E is the original edge dataset of the graph G. The i th node v_i in the node dataset V has a node feature representation of $v_i=\left[x_i^1,x_i^2,x_i^3,...,x_i^N\right]$, where N is the number of node features.

In this paper, node similarity is evaluated by calculating the cosine distance between nodes. For any two points v_i,v_j in graph G, the similarity $sim_{i,j}$ has:

$$sim_{i,j}=\cos<v_i,v_j>=\frac{<v_i,v_j>}{|v_i||v_j|}=\frac{\Sigma_{k=1}^{N}(x_i^k x_j^k)}{\sqrt{\Sigma_{k=1}^{N}x_i^{k^2}}\cdot\sqrt{\Sigma_{k=1}^{N}x_j^{k^2}}} \tag{7}$$

For the computed cosine similarity set, the algorithm uses K-Nearest Neighbors (KNN) to construct the graph structure under the feature similarity perspective as one of the inputs, as:

$$\hat{E}=KNN(k,Sim) \tag{8}$$

where $\hat{E}$ is the neighbor adjacency matrix in similarity perspective constructed based on similarity, and Sim is the similarity matrix between nodes.

(2) Fraud detection based on multi-view graph embedding

In this paper, we make full use of the original viewpoint information of the graph network and the similarity viewpoint information to propose a fraud detection algorithm based on multi-view graph neural network. It mainly contains a feature extraction module based on the multi-view neighbor matrix of graph data, a music emotion analysis module for live network broadcasts, and an embedded information fusion module based on the attention mechanism.

In this paper, the node feature embedding module consisting of two layers of graph convolution is constructed for the two perspectives of original critical matrix and similarity critical matrix respectively. The graph structure is utilized to aggregate the neighbor information of nodes through graph convolution

v_i of a node x_i, the set of its neighboring nodes is $N(v_i)$, and the neighboring nodes' feature representation is $X_{N(v_i)}$. Then the new feature representation h_i of the node v_i can be computed by:

$$h_i = \sigma\left(\sum_{j \in N(v_i) \cup \{i\}} \frac{1}{c_{ij}} W x_j \right) \tag{9}$$

where W is the learnable parameter matrix, σ is the activation function, and c_{ij} is the normalization coefficient for balancing the degrees of different nodes.

Through multi-layer graph convolution operation, GCN is able to gradually aggregate the information of more distant neighboring nodes. Assuming there are K layers of graph convolution layers, the output of the k th layer can be expressed as:

$$H^{(k)} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(k-1)} W^{(k)}) \tag{10}$$

$$H^{(0)} = X$$

where $W^{(k)}$ is the parameter matrix of the k th graph convolutional layer, $\hat{A} = A + I$ is the result of the adjacency matrix A plus the self-connection, $\hat{D}$ is the diagonal angle matrix, and σ is the activation function.

In order to effectively aggregate the embedding vector information obtained by the nodes based on the neighbor matrices of different viewpoints, this paper adopts the attention mechanism. In the algorithm proposed in this paper, the model uses the attention mechanism to aggregate the embedding vectors originating from different viewpoints. For the set of node feature vectors obtained by the graph convolution module $Z = [z_1, z_2, ..., z_n]$, and n is the number of viewpoints of the dataset introduced by the model.

The model introduces a weight matrix W, computes and maintains the weight matrix through the attention mechanism, and computes the weighted fusion of the embedded vectors using the attention distribution to obtain the weighted vector $\hat{z}$. i.e.,:

$$\hat{z} = W \cdot Z = \Sigma w_i \cdot z_i \tag{12}$$

In the algorithm proposed in this paper, the model uses the attention mechanism to aggregate the embedding vectors originating from different embeddings.

For the set of node feature vectors Z obtained by the graph convolution module, the feed-forward neural network layer is used to compute the attention weights, and its output is the attention score vector $\omega = [w_1, w_2, ..., w_n]$, where w_i denotes the attention score of the corresponding node z_i. The attention score can be calculated using Eq:

$$w_i = f(z_i) = Linear(z_i) \tag{13}$$

where $f(\cdot)$ denotes the feed-forward neural network layer.

The attention scores need to go through Softmax operation to get the attention distribution to ensure the normalization of the attention weights.The Softmax process is shown in Eq:

$$\alpha_i = Softmax(w_i) = \frac{\exp(w_i)}{\sum_{j=1}^{n} \exp(w_j)} \tag{14}$$

where α_i denotes the attention weight of node z_i.

Finally, the node feature vectors are weighted and fused by the attention weights to obtain the weighted vector $\hat{z}$. The weighted fusion operation can be realized using the following equation:

$$\hat{z} = \sum_{i=1}^{n} \alpha_i z_i \tag{15}$$

where $\hat{z}$ denotes the weighted node feature vector.

For the node embedding vectors fused by the attention module, this paper employs a multilayer perceptron (MLP) to map the fused node representations into a high-dimensional space and ultimately to obtain the prediction confidence of the model.The input layer of the MLP receives the fused embedding vectors $\hat{z}$ from the attention module and passes them to the first hidden layer. Each neuron in the hidden layer performs the following operations:

$$h_j^{(1)} = \sigma\left(\sum_{i=1}^{n} w_{ij}^{(1)} \hat{z}_i + b_j^{(1)}\right) \tag{16}$$

where $h_j^{(1)}$ is the output of the j th neuron in the first hidden layer, $w_{ij}^{(1)}$ is the weight connecting the input layer and the first hidden layer $b_j^{(1)}$ is the bias of the j th neuron in the first hidden layer, and σ is the activation function.

The output of the hidden layer is used as the input of the next layer, and the output of the output layer is finally obtained after multiple hidden layers. The output of the output layer $\hat{y}$ represents the prediction result of the model, which is calculated as:

$$\hat{y} = \sigma\left(\sum_{j=1}^{m} w_{kj}^{(2)} h_j^{(2)} + b_k^{(2)}\right) \tag{17}$$

where $h_j^{(2)}$ is the output of the j th neuron in the last hidden layer, $w_{kj}^{(2)}$ is the weight connecting the last hidden layer and the output layer, and $b_k^{(2)}$ is the bias of the k th neuron in the output layer.

With this approach, the complex relationship between multiple perspectives of information can be captured more accurately and the fraudulent nodes can be effectively identified and understood in a higher dimensional space, which improves the detection of anomalies, and provides a more powerful tool for further in-depth understanding of the behavioral patterns in complex networks.

4. Detection and Analysis of Fraud in Webcasting

4.1. Fraud Detection Model Performance Experiments

Among all the experiments, 20 runs were conducted in this paper and the average value was taken as the final result. The training set, validation set and test set are 35%, 30% and 35% respectively.

The hyperparameters of the model in this paper are set as follows: the learning rate is 0.001, the Adam optimizer is used, the weight decay is set to 0.003, and the hidden layer dimension of the node features is 128.

The Yelp dataset is derived from Yelp, the largest online review site in the U.S., which focuses on user reviews of hotels and restaurants.The Dropout rate on the Yelp dataset is 0.2, and that on the Amazon dataset is 0.3, and the number of training cycles is 200.

The Amazon dataset, on the other hand, focuses on user reviews of musical instrument products, which are also labeled as normal or fraudulent, and originates from Amazon, the largest online retail platform in the U.S. In the Amazon dataset, the batch size is set to 512. In the Yelp dataset, the batch size is set to 512.

4.1.1. Experimental Setup

In order to validate the effectiveness of the models on fraud detection datasets, some of the most representative baseline models are selected for comparison in this chapter, which are:

GCN: a GNN model implemented through spectral domain convolution.GCN captures spatial relationships in the graph structure by utilizing the spectral theory of graphs for feature aggregation of nodes.

GAT: a null-domain GNN model that uses an attention mechanism in aggregating neighboring nodes.GAT effectively captures the importance and interrelationships among nodes by assigning different attention weights to each node.

GraphSAGE: a model that solves the memory explosion problem through neighborhood sampling.GraphSAGE is able to generate node embeddings from the local neighborhoods of nodes, thus enabling efficient processing of large-scale graph data.

GraphConsis: a heterogeneous graph model that performs balanced sampling optimization by measuring the distance of neighboring nodes, which is mainly applied to risk control scenarios.GraphConsis improves the performance and stability of the model in processing heterogeneous graph data by optimizing the sampling process.

CARE-GCN: an end-to-end model that focuses on solving the problem of node camouflage by fraudsters.CARE-GCN effectively improves the accuracy of fraud detection by identifying the similarities between fraudulent nodes and normal nodes.

DR-GCN: A GNN specifically designed to deal with the problem of category imbalance.DR-GCN enhances the model's ability to recognize nodes of a few categories by adjusting the weight assignment during graph convolution.

GPRGNN: a model that can optimize topological information extraction and node characterization through GPR architecture and is able to learn adaptive GPR weights.GPRGNN improves information aggregation by fine-tuning the relationship weights between nodes.

FAGCN: a frequency domain model that proposes convolution kernels for extracting high-frequency and low-frequency information separately.FAGCN improves the accuracy of feature extraction and the generalization ability of the model by differentiating between high-frequency and low-frequency information in the processing graph.

PC-GNN: a fraud detection model that solves the sample imbalance problem by undersampling and uses node distance measurements to resample neighboring nodes.

This chapter uses three metrics commonly used in node classification: the F1, AUC, and Gmean. Where Gmean's identification of minority classes in extremely unbalanced datasets is critical to improving the utility of the model.Gmean ensures that the model must perform well on both majority and minority classes to achieve a high score by taking a geometric mean of the recall.

4.1.2. Analysis of Experimental Results

(1) Model comparison analysis

The F1 scores of different models on the two datasets are shown in Fig. 2, and the resultant data from the experiments show that compared with GCN, GAT, and GraphSAGE, this paper's model has an absolute advantage in all the F1 indicators. This paper's model outperforms GCN, GAT, and GraphSAGE by 0.1298, 0.2186, and 0.2364 on the dataset Yelp, respectively.

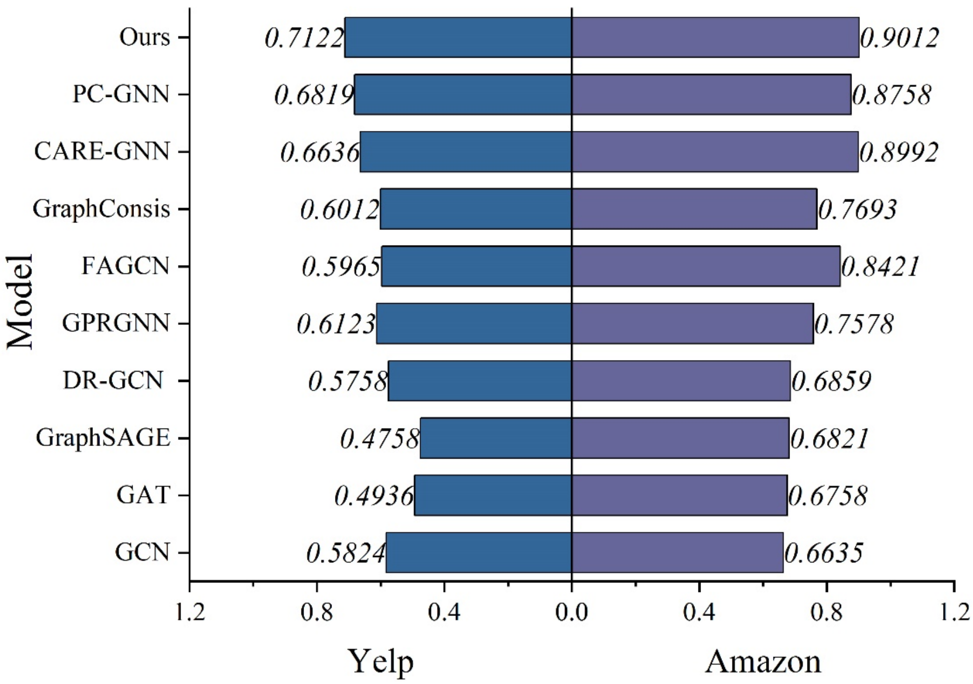

Figure 2. F1 scores on two datasets different models.

The AUC scores of the different models on the two datasets are shown in Fig. 3, where the model of this paper outperforms the competing model PC-GNN by 0.0275 and 0.0091 on the dataset Yelp and dataset Amazon, respectively.

Comparing the performance of AUC scores of the models on the two datasets, it can be observed that the models perform more prominently on the dataset Amazon.

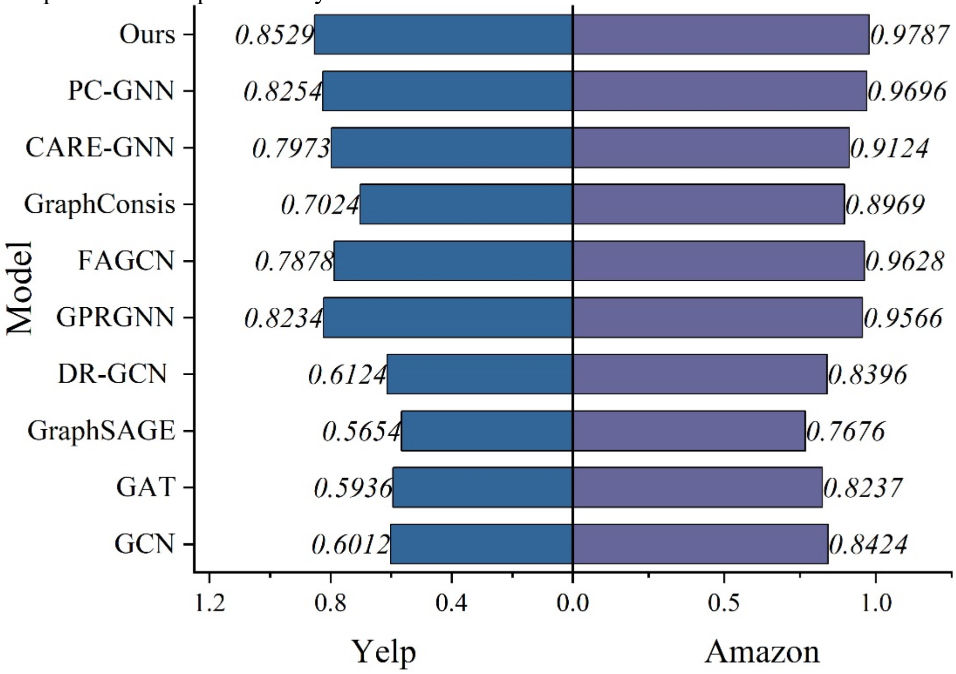

Figure 3. AUC scores on two data sets.

The Gmean scores of the different models on the two datasets are shown in Fig. 4, and the combined Gmean scores of the models show that the model in this paper has the highest Gmean score. It is 0.7682 and 0.9341 on dataset Yelp and dataset Amazon respectively.

Model GAT performs poorly on dataset Yelp, with a Gmean score of only 0.1896. The Gmean scores of each model on dataset Amazon are higher than 50%, indicating that each model performs well on both the majority class and the minority class in dataset Amazon.

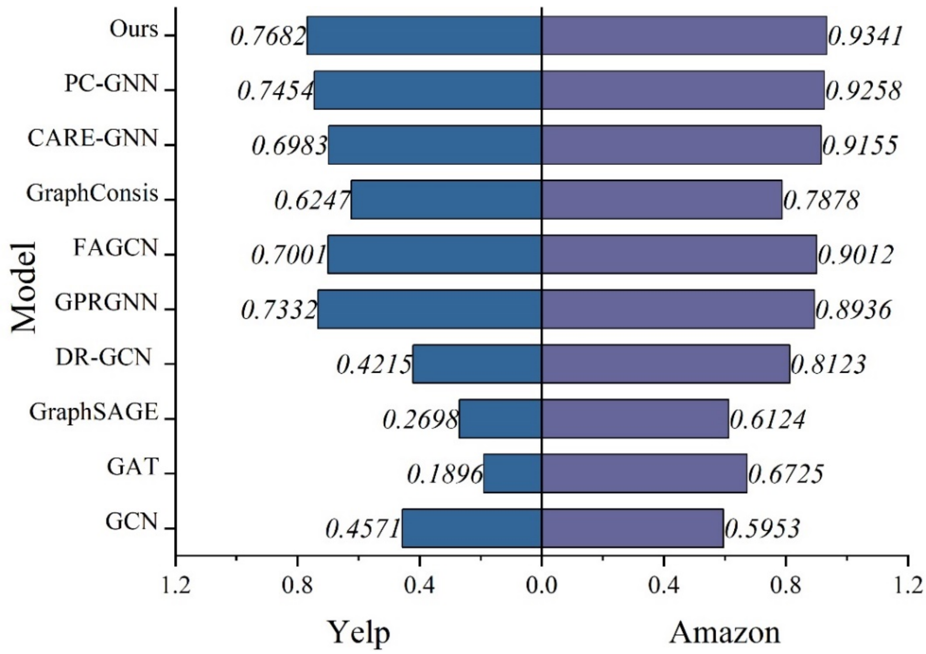

Figure 4. Different models score Gmean scores on two data sets.

(2) Analysis of Over-smoothing Handling Ability

In order to have a more intuitive feeling of the processing ability of this paper's model for oversmoothing, this chapter visualizes the dimensionality reduction of the original features and the embedded representation of the model on the dataset Yelp.The initial features of the Yelp dataset are shown in Fig. 5.

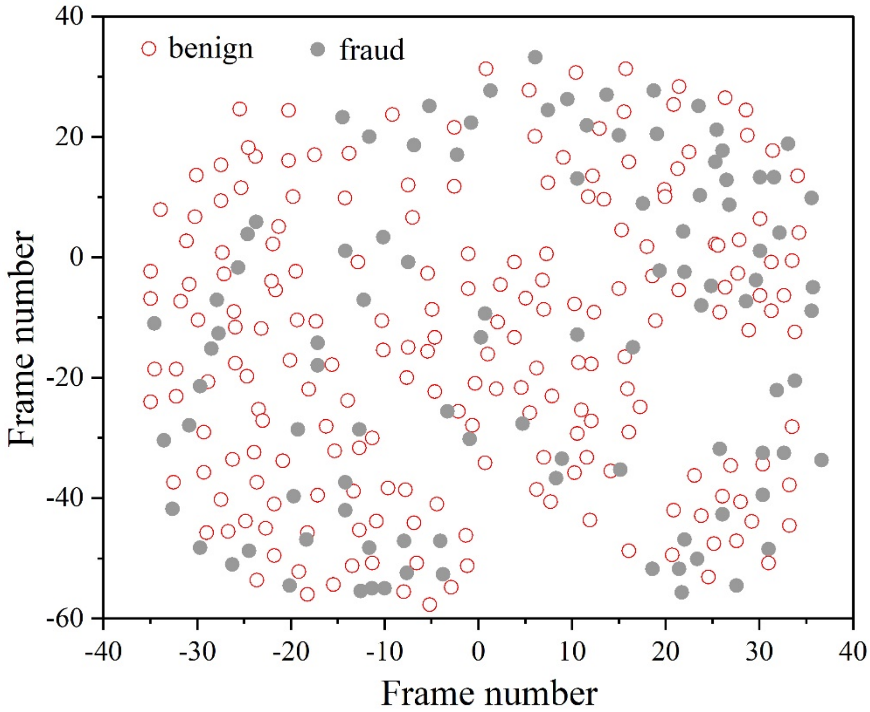

Figure 5. The initial characteristics of the the active data set.

The embedding vector dimensionality reduction visualization of the model in this paper is shown in Figure 6. On the Yelp dataset, the benign and malignant nodes are not distinguished after the original features are downscaled. While the final layer embedding of this paper's model distinguishes most of the benign nodes from the malignant nodes, proving that the attention enhancement module has a stronger ability to solve the oversmoothing problem.

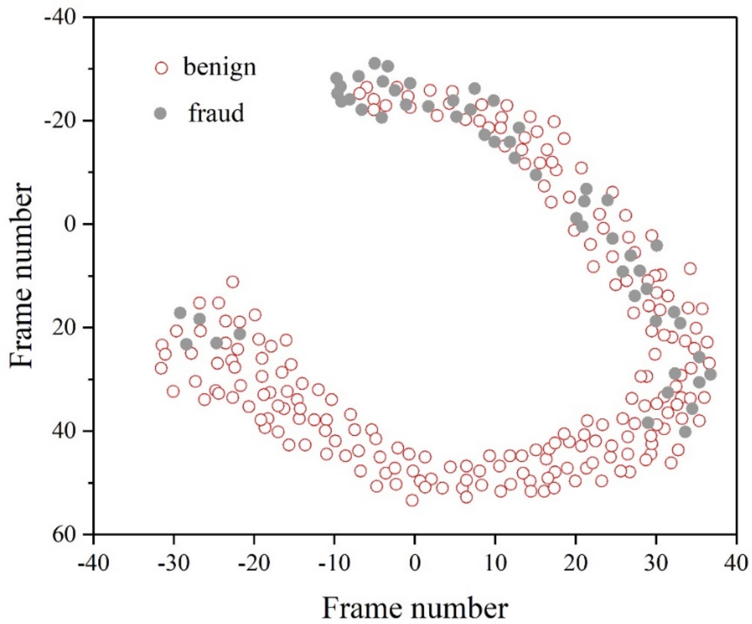

Figure 6. The model's embedded vector degradation visualization.

4.2. Classification and Detection of Web Host Fraudulent Behavior

4.2.1. Experimental Setup

The preprocessed live broadcast interaction behavior dataset of online anchors in the previous article is used as the anchor fraud identification data, and the labels of the dataset are set to 0 to 4, where 0 represents normal interactive discourse, 1 to 4 represents four types of fraud, 1 represents "irrational consumption" fraud, 2 represents "inferior product" fraud, 3 represents "false information" fraud, and 4 represents "misconception" fraud. When the training set is unbalanced, set the ratio to 1:1:1:1:1.

4.2.2. Experimental Results and Analysis

The experiments compare the five-classification detection effect of this paper's model under different sampling methods. The effect of sampling methods on the model detection results is shown in Figure 7. Focusing first on Macro-R (arithmetic mean recall), compared with integrated sampling, the Macro-R value is significantly higher in the case of using SMOTE sampling method, both for the baseline model GCN and the proposed model in this paper. The Macro-R value obtained by this paper's model using the SMOTE sampling method is 0.9986. The Macro-P (arithmetic mean precision) and Macro-F1 values are relatively lower, but this paper's model is still ahead of the other compared methods.

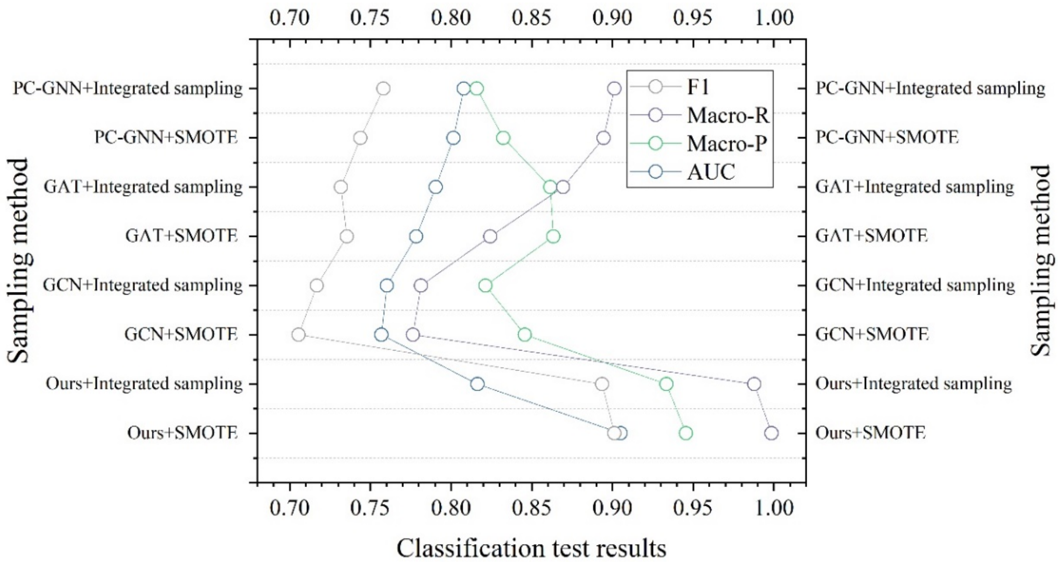

Figure 7. The sampling method affects the results of the model.

The confusion matrix for the five-classification detection experiment based on the model in this paper is shown in Figure 8.

For the majority class of normal interactions, the model correctly detected 7812 comments and incorrectly detected 83 comments, with a recall rate of 98.95%. The number of normal comments detected by the model is 7880, but the actual number of normal comments is 7812, with a precision rate of 99.14%.

Similarly, for the first type of fraud “unhealthy food” fraud, the second type of fraud, the third type of fraud, and the fourth type of fraud, the precision rate is calculated to be 96.27%, 96.65%, 90.20%, and 92.67%, in that order. The arithmetic mean accuracy of the model is therefore Macro-P is 93.95%.

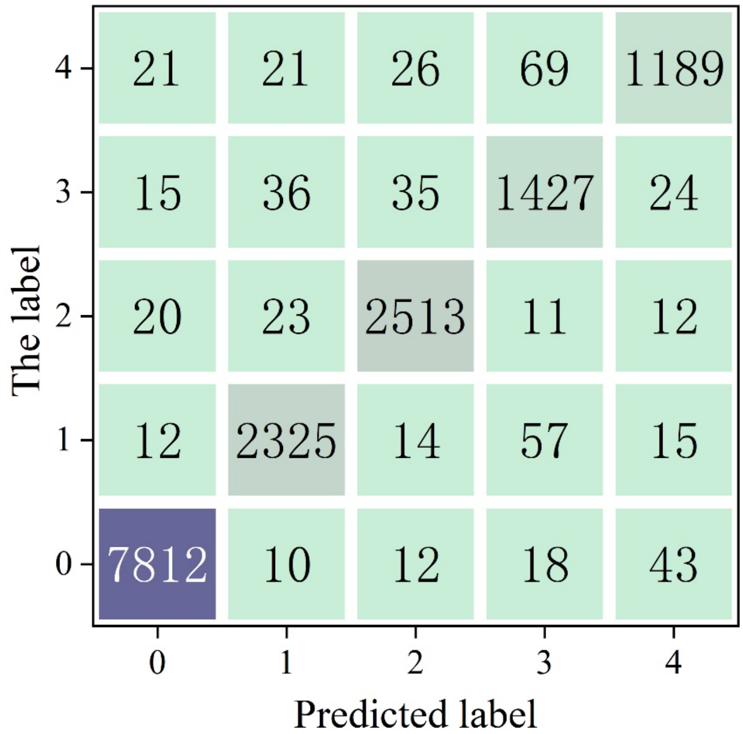

Figure 8. The confusion matrix of the five classification test experiment of the model.

The five-classification detection results of each model are shown in Fig. 9. From the performance detection results of each model in the figure, it can be seen that the model of this paper has outstanding performance in F1, Macro-R, Macro-P, and AUC metrics. This paper's model outperforms the GCN, GAT, and GraphSAGE models by 0.0256, 0.0479, and 0.1 on the F1 metrics, respectively.

Overall, all the metrics of this paper's model are higher than the other models, indicating that in the task of five-classification detection, compared with the other models, this paper's model does find more feature information for a few classes, has better detection ability, and is able to predict and recognize fraudulent behaviors in webcasting.

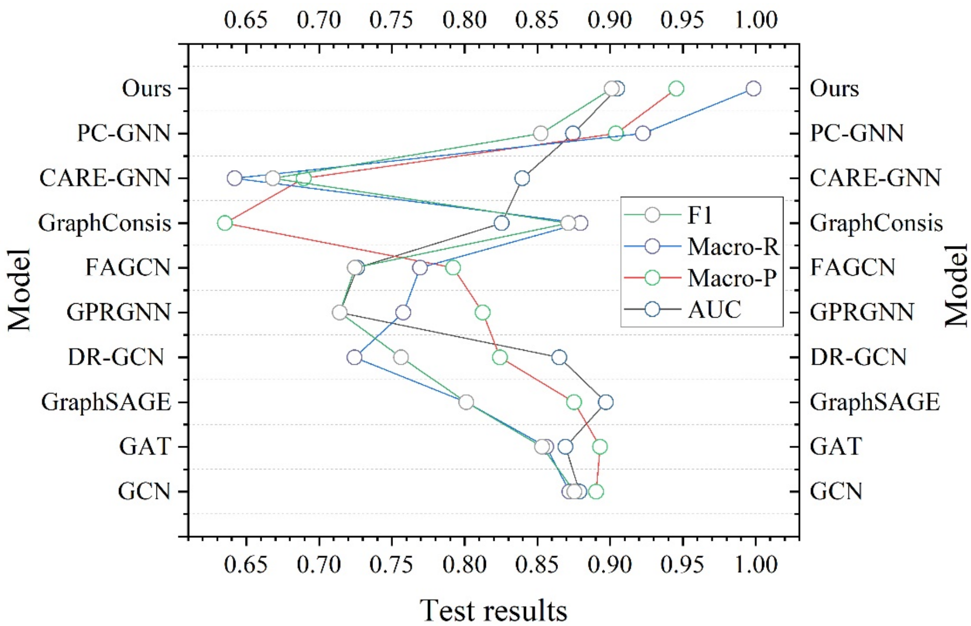

Figure 9. Five classification test results.

5. Conclusion

In this study, the key technical difficulties in the detection of webcasting bounty fraud are effectively solved by constructing a graph neural network model that integrates music sentiment analysis and user interaction time sequence. The model shows significant advantages in dealing with multimodal information fusion and identification of complex interactions, verifying the application value of graph neural network technology in the field of fraud detection.

The experimental results fully demonstrate the effectiveness of the proposed method. In the standard dataset test, the model achieves significant performance improvement compared to the traditional method, and the AUC metric is improved by 0.0275 and 0.0091 compared to the competing model PC-GNN on the Yelp and Amazon datasets, respectively. In the webcaster fraud detection task, the model shows excellent classification performance, with a recall rate of 98.95% for normal interaction behavior, a precision rate of 99.14%, and an arithmetic average precision rate of 93.95%. The model successfully identifies four types of typical fraudulent behavior patterns, providing accurate risk warning capabilities for live broadcast platforms.

In terms of technological innovation, the music sentiment analysis module effectively captures the sentiment features of live audio through the Mel frequency cepstrum coefficient, the multi-view graph neural network architecture realizes the deep fusion of different dimensional information, and the attention mechanism further optimizes the feature representation learning process. The model provides an important technical guarantee for the healthy development of the webcasting industry, helps to build a safer and more trustworthy digital entertainment environment, and protects the legitimate rights and interests of consumers.

References

1. Zou, S. (2018). Producing value out of the invaluable: A critical/cultural perspective on the live streaming industry in China. TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society, 16(2), 805-819.
2. Liu, Z. (2020, August). Research on the current situation and future trend of web celebrity e-commerce live streaming industry. In 2020 The 4th International Conference on Business and Information Management (pp. 133-137).
3. Lu, Y. (2024). Identification of the Legal Nature of the Behavior of Webcasting Reward and Analysis of the Nature of the Fraud of Inducing Reward Class. Lecture Notes in Education Psychology and Public Media, 73, 40-46.
4. Xiong, Y., & Chen, B. (2025). Evolutionary strategies in live-streaming e-commerce: A four-player game model analysis. Human Systems Management, 01672533251317071.
5. Bakri, N. (2024). REVIEW OF THE CONCEPT OF JU'ALAH ON INCOME EARNED FROM LIVE STREAMING ON SOCIAL MEDIA: A CASE STUDY ON TIKTOK APPLICATION. JURISTA: Jurnal Hukum dan Keadilan, 8(2), 762-783.
6. Li, H., Wang, Z., Yuan, Z., & Yan, X. (2023). Multidimensional Evaluation of Consumers' Shopping Risks under Live-Streaming Commerce. Sustainability, 15(19), 14060.
7. Liu, T., Tan, C. K., Yang, X., & Li, M. (2023). Zhibo gonghui: China's 'live-streaming guilds' of manipulation experts. Information, Communication & Society, 26(6), 1210-1225.
8. Xu, Y., Qi, J., Kong, J., & Zhang, W. (2024). Strategic decision making in live streaming e-commerce through tripartite evolutionary game analysis. Plos one, 19(7), e0305427.
9. Rafique, M. Z., Van Goethem, T., Joosen, W., Huygens, C., & Nikiforakis, N. (2016, February). It's free for a reason: Exploring the ecosystem of free live streaming services. In Proceedings of the 23rd Network and Distributed System Security Symposium (NDSS 2016) (pp. 1-15). Internet Society.
10. Hwang, S. H., Kim, G. Y., Myeong, S. H., Yun, T. S., Yoon, S. M., Kim, T. H., & Euom, I. C. (2022). Vulnerabilities of Live-Streaming Services in Korea. Sensors, 22(10), 3766.
11. Wu, Q., Sang, Y., Wang, D., & Lu, Z. (2023). Malicious selling strategies in livestream e-commerce: A case study of Alibaba's Taobao and ByteDance's TikTok. ACM Transactions on Computer-Human Interaction, 30(3), 1-29.
12. Yang, K. (2024). How to prevent deception: A study of digital deception in "visual poverty" livestream. New Media & Society, 14614448241285443.
13. Watters, P. (2024). Exposing the Dark Side: Scams and Cybersecurity Risks in Indonesia's Illicit Sports Streaming Scene. Available at SSRN 4954969.
14. Wu, Y., Niu, G., Chen, Z., & Zhang, D. (2022). Purchasing social attention by tipping: Materialism predicts online tipping in live-streaming platform through self-enhancement motive. Journal of Consumer Behaviour, 21(3), 468-480.
15. Ye, F., Ji, L., Ning, Y., & Li, Y. (2024). Influencer selection and strategic analysis for live streaming selling. Journal of Retailing and Consumer Services, 77, 103673.
16. Deng, Y., He, C., Zou, Y., & Li, B. (2025, April). " Auntie, Please Don't Fall for Those Smooth Talkers": How Chinese Younger Family Members Safeguard Seniors from Online Fraud. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

17. Saputra[1], D. D., Jamaludin, A., Nurhasanah, I. Y., & Maulana[4], I. (2024). Policy Crisis in Addressing Anonymity in Live Streaming. In Proceedings of the ASEAN Conference on Sexual Exploitation of Children (ACOSEC 2024). (p. 7). Springer Nature.
18. Luo, L., Hanyi, S., Zheng, Y., & Yuan, Y. (2025). To Gift or Not: Understanding Gifting Behavior on Live Streaming Platforms from the Perspective of Social Influence and Herding. International Journal of Human–Computer Interaction, 41(10), 5831-5848.
19. Zhang, Z., & Liu, F. (2024). Gift-giving intentions in pan-entertainment live streaming: Based on social exchange theory. Plos one, 19(1), e0296908.
20. Fan, J., Peng, L., Chen, T., & Cong, G. (2024). Regulation strategy for behavioral integrity of live streamers: From the perspective of the platform based on evolutionary game in China. Electronic Markets, 34(1), 21.
21. Li, Z., Wang, H., Zhang, P., Hui, P., Huang, J., Liao, J., ... & Bu, J. (2021, August). Live-streaming fraud detection: A heterogeneous graph neural network approach. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (pp. 3670-3678).
22. Baratzadeh, F., & Hasheminejad, S. M. (2022). Customer Behavior Analysis to Improve Detection of Fraudulent Transactions using Deep Learning. Journal of AI and Data Mining, 10(1), 87-101.
23. PK, R., Khaparde, A., Bendre, V., & Katti, J. (2025). Fraud detection and prevention by face recognition with and without mask for banking application. Multimedia Tools and Applications, 84(2), 781-804.
24. Meryam Telmem,Naouar Laaidi & Hassan Satori. (2025). The impact of MFCC, spectrogram, and Mel-Spectrogram on deep learning models for Amazigh speech recognition system. International Journal of Speech Technology,28(1),1-14.
25. Vincenzo Capone,Angelo Casolaro & Francesco Camastra. (2025). Spatio-temporal prediction using graph neural networks: A survey. Neurocomputing,643,130400-130400.
26. Luyi Bai,Linshuo Xu & Lin Zhu. (2025). Question answering over temporal knowledge graphs based on time sensitive graph neural network. Applied Soft Computing,177,113257-113257.