DIEET: Knowledge–Infused Event Tracking in social media based on Deep Learning

Jun Ge
Suqian University
LEILEI SHI
Jiangsu University
Lu Liu
l.liu@leicester.ac.uk

University of Leicester
Zi-xuan Han
Jiangsu University
Anthony Miller
University of Leicester

Research Article

Keywords: Human behavior analysis, Event tracking, Deep neural network, Diffusion behavior, Interest evolution behavior

Posted Date: August 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3240718/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Peer-to-Peer Networking and Applications on April 17th, 2024. See the published version at https://doi.org/10.1007/s12083-024-01677-z.

DIEET: Knowledge–Infused Event Tracking in social media based on Deep Learning

Jun Ge[1,5], Lei-lei Shi[2,3], Lu Liu[4], Zi-xuan Han[2,3], Anthony Miller[4]
[1]School of Information Engineering, Suqian University, Suqian, China
[2]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China
[3]Jiangsu Key Laboratory of Security Technology for Industrial Cyberspace, Jiangsu University, Zhenjiang, China
[4]School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK
[5]Jiangsu Province Engineering Research Center of Smart Poultry Farming and Intelligent Equipment, Suqian, China
Corresponding author: Lu Liu, Email: l.liu@leicester.ac.uk；Lei-lei Shi, Email: 937109472@qq.com

Abstract: The rapid expansion of the mobile Internet has led to online social networks becoming an increasingly integral part of our daily lives, this offers a new perspective in the study of human behavior. User' interest groups are now considered crucial elements within social networks. The objective of event evolution and tracking within social networks is the ability to monitor the real-time evolution of user' interests based on the previous diffusion behavior of influence disseminators and to anticipate future diffusion behavior of users. However, prior studies have only addressed the evolution regarding a single interest or the evolution of user interest in the entire interest space, ultimately with no modeling of user interest evolution within event tracking. In order to address these challenges, this study proposes a knowledge-infused deep learning-based event tracking model named DIEET (Diffusion and Interest Evolution behavior modeling for Event Tracking). This model accurately predicts the propagation and interest evolution behavior in event tracking by considering both propagation and interest evolution behavior. Specifically, the DIEET model incorporates the interval time, the number of times, the sequence interval time, and finally user preference for the event of interest, greatly improving the accuracy and efficiency of event evolution prediction. The experiments conducted on real Twitter datasets detail the proposed DIEET models' ability to greatly improve the tracking of the state of user interest alongside the popularity of event propagation, and DIEET also has superior prediction performance compared to state-of-the-art models in terms of identifying user dynamic interest. Therefore, the aforementioned model offers promising potential in the ability for predicting and tracking the evolution of user interest and event propagation behavior on online social networks.

Keywords: Human behavior analysis; Event tracking; Deep neural network; Diffusion behavior; Interest evolution behavior;

I. INTRODUCTION

The increasing role of social media within people's daily lives has ultimately expanded the scope of human behavior analysis [1-4,46]. Advanced machine learning algorithms such as deep learning [8] have shown great potential in analyzing social media data [5-7]. Dynamic heterogeneous social networks, including event tracking alongside other types of analysis, have recently gained considerable recognition [6-9]. Scholars from various fields have examined online social networks from various differing angles in order to identify events and evaluate the overall network [10-13]. This topic continues to be a popular methodology and a challenging research topic [14].

Social network event tracking technology refers to the process of identifying the publisher's identity and also restoring information dissemination by providing appropriate diffusion resources for users based on their range of interests [15]. User interests and hobbies are potentially influenced by their learning stage and also their cognitive ability, which changes frequently during the event diffusion process. Event tracking tasks include tracking users' real-time interests through their diffusion history and also the prediction of their performance within future diffusion based on their interests.

Although researchers and related institutions have made significant progress in social network topic identification [16-18] and tracking technology of social network information [19,20], research on event traceability is recognised as still relatively limited. Most research on event tracking and traceability technology in social networks focuses on microblog topics, this involves tracking the subsequent topic development utilising known hot topics, rather than backtracking the propagation path of a specific microblog. In this paper, we relate to the research results of microblogging topics and also relevant knowledge of topic detection alongside tracking (TDT) technology to complete the research regarding the tracking of hot events within social networks.

However, existing event tracking methods disregard the impact of user interest evolution behavior during the diffusion process. Within the field of diffusion psychology, many scholars have recognized the evolution behavior of human interest, as proposed by the Ebbinghaus interest evolution curve theory [21], this indicates that user interest will eventually evolve and the impact of interest evolution is ultimately declining interest satisfaction. The number of times users repeat transmission of interests and also the time interval from the last transmission of potentially affect interest evolution. However, these studies only consider some information in the effect of the evolution of user interests, while ultimately ignoring the impact of user's original interests on the evolution of interests. Additionally, previous studies focusing on the evolution of interest, only took in account the evolution of a single interest, therefore ignoring the multi-interest evolution of users [22-24].

In order to address these issues, this paper proposes a knowledge-infused deep event tracking model named DIEET (Diffusion and Interest Evolution behavior modeling for Event Tracking), which considers both the influence diffusion and interest evolution of influential spreaders. The DIEET model fits influential spreaders' diffusion and interest evolution behavior, updates, alongside outputs regarding their interest satisfaction degree in real-time, and ultimately predicts their future performance.

(1) In order to achieve this, the DIEET model considers four factors that affect the evolution of influential spreaders' interest based on diffusion psychology: Firstly the interval time of users' repeated dissemination of interests, followed by the number of repeated dissemination of interests, then

the interval time of sequential dissemination, and finally the degree of users' satisfaction of interests. Our model fits the change of interest satisfaction degree caused by users' interest evolution behavior;

(2) Based on deep neural network technology, the DIEET model designs an event tracking neural network based on RNN and memory neural network. The network includes an attention layer, interest evolution layer, propagation layer, prediction layer, and interest output layer. Within the interest evolution layer, there lies a fully connected network which calculates memory erasure vector and also memory update vector in order to fit the interest evolution behavior. In the diffusion layer, LSTM network is utilized, the results of users' choices at the end of diffusion are utilised as indirect feedback of interest satisfaction degree. According to the memory erasure vector and memory update vector, the intermediate embedding of interest satisfaction degree after interest evolution is obtained, which is then used as the input of the diffusion layer in order to obtain the interest satisfaction degree embedding of interest evolution and diffusion;

(3) Experimental results on real Twitter datasets show that the DIEET model can potentially effectively model influential spreaders' propagation behavior and also interest evolution process, alongside tracking users' interests in real-time, and also has shown better prediction performance than other existing models.

The remainder of the paper describes related work on event evolution and tracking in section II, defines relevant ideas and symbols, and also poses questions in section III. The suggested preprocessing approach is described in-depth in section IV, while section V presents the findings and analysis of the DIEET comparison experiment. Ultimately the final section provides our conclusions and outlines plans for future research.

II. RELATED WORK

In recent years, research on social media has generated significant attention [25,26]. The study of the evolution mechanism of hot event information in social networks is a research topic which contains important scientific and practical value. An event evolution model has been identified as a crucial component of hot event detection and tracking technology in social networks. A good event evolution model has the ability to improve the accuracy and efficiency of hot event recognition, therefore enhancing the performance of hot event detection and also tracking technology. This research involves many issues, these include network structure analysis, text content analysis, and also processing, our research also generates extensive interest among researchers within complex network, information retrieval, and also big data fields.

Since the emergence of social media, numerous problems have arisen, including a vast amount of user-generated content and also information overload. Consequently, the evolution of hot events within social networks has generated huge attention. By tracking the evolution process of popular events and the evolution of users' views and attitudes, government departments can potentially gather and also analyze network security information, thus allowing them to take timely and appropriate control and guidance measures [27,28].

Event evolution pertains to the propagation and diffusion process of hot events within social networks, this is usually associated with the form and mode of hot event propagation. For instance, captured on video sharing websites, the evolution of video events is typically measured by the number of views and shares [29], while the dynamic communication of events in news platforms is represented by the number of news comments [30]. The evolution of social network event information can potentially be presented in various forms, this can be summarized into two perspectives: the scope of communication [31] and also the cycle of communication [32]. Regarding the scope of dissemination this measures the dynamic dissemination of information from a spatial perspective, thus on the overall trend of information dissemination. Communication cycle measures the dynamic dissemination of information from a temporal perspective, emphasizing the speed and duration of information dissemination in the network.

Traditional event evolution research mainly focuses on high-frequency keyword detection [33] and also clustering [34], or alternatively detection based on topic models [35]. Ultimately, most of these methods utilise some high-frequency keywords as an overview of emergencies or alternatively hot events. However, hot events in real life evolve over time, ultimately making it crucial to track their evolution and development. Additionally, the evolution of hot events is also related to changes in the interests of influential communicators during the event evolution process. Therefore, it is essential to pay attention to both hot events and also the evolution interests of influential spreaders under such events, rather than just focusing on hot events alone.

Currently, various research studies have been conducted regarding event evolution. The event link detection (SLD) model, proposed in [36], is designated as an essential method in traditional topic detection and tracking research, which ultimately aims to discover the relationship between documents. Identified within [37] is the construction of a multi-vector event model in order to describe events and also calculates the similarity between vectors in order to describe the relationship between events, alas this does not distinguish different weights of words, geographical location, name, and also background information. .Research witihn [38] proposes an event evolution ranking method, this evaluates the possible relationships within the process of event evolution by integrating the similarity of event content, background information, and also document distribution without considering other event attributes such as geographic location and also participants. Events within [39] consider different combinations of multiple feature information, alas these methods are applied to news long text corpus and cannot be directly applied to microblog short text.

To address the aforementioned social network microblog short text event evolution problem, investigations contained within [40] proposes a method to construct the event talker model in the microblog network that can dynamically detect the core subgraph of the event by using the time sliding window and also monitor the evolution process of the subgraph. Proposed research within [41] uses a graph optimization model and also dynamic pseudo relevance feedback to obtain the evolution process of related microblogs and events; however, this method requires users to query keywords, which limits its effectiveness for short text event evolution without prior knowledge. Therefore, analysis contained within [42] proposes a time hashtag late Dirichlet allocation (TH-LDA) model based on microblog time and also a tag to better detect hot events and also to identify unmarked microblogs that belong to the same hot event as tag text. Moreover, research within [42] proposes a hot topic life cycle model (HTLCM model) to continuously track the evolution trend of hot event topics .

In addition, information overload has become an increasingly serious anomoly in the evolution of social network hot events, therefore making analysis more difficult in order to find useful information about hot events. Therefore, in order to quickly find the relevant information related to hot events in unstructured, diverse, and dynamic social networks have become a hot topic within information retrieval, data mining, social networks, and also various other fields. Within social networks, event evolution based on user interest evolution mainly refers to the search for people and also information related to specific user interest or particular event alongside its change process. There consists of two main solutions, which are collaborative filtering-based method and interest community discovery-based method.

The collaborative filtering method finds similar users of the target user by analyzing the score matrix of users and also social information, which predicts the interest of similar users to ultimately determine whether to recommend the discovered events for the current user according to the predicted score level. However, this method has two problems: data sparsity and also cold start. The interest-based community discovery method finds people or information related to users through the community, this is originated from the idea of "birds of a feather flock together." However, most existing interest community discovery algorithms cannot accurately reflect the real network due to the nature of social networks becoming dynamic with users' interests and also changing over time.

As mentioned within the above study, while considerable progress has been achieved in the research and application of hot event evolution in social networks, the evolution model itself still has a number of flaws that impair the overall effectiveness of hot event detection and tracking technologies:

(1) The existing event evolution model lacks the ability to identify the user's interest community in hot events, and also cannot identify the key microblogs and communities The change process regarding the influence of key people lacks the tracking and also identification ability of microblog dynamic influence and user dynamic influence.
(2) It is difficult to trace the evolution process of hot events according to influential spreaders interest migration, and also to cluster and trace hot events automatically and accurately.
(3) In the process of event evolution, the low recognition rate of new and old hot events leads to the low control ability of event evolution, and it is difficult to track the evolution of hot events and the interest evolution process of influence disseminators efficiently and accurately.

III. PROBLEM DEFINITION

Here, we define U as the user set, I as the interest set, E as the event set. Each user spreads independently without affecting each other. The diffusion history of users is $H_u = [(e_1, d_1), (e_2, d_2), ..., (e_t, d_t)]$, where: e_t is the event diffusion done by the user at t time; d_t is the result of the diffusion, and d_t =1 means the diffusion is correct, $d_t = 0$ indicates the wrong diffusion. I_t, belongs to I is the set of interest involved in diffusion set. Matrix M^I ($d_i \times |I|$) is the embedded representation of $|I|$ interest and hobbies in the whole interest space, among which, the d_i dimension vector is an embedded representation of interest. Matrix $M_{t-1}{}^{v}$ ($d_v \times |I|$) is embedded in the interest space of t-1. The user's satisfaction of each interest at the end of time transmission is represented by $|I|$ dimension vector $value_{t-1}$, where the value of each dimension of the vector is between (0,1): the value is (0,1): the value is 0, indicating that the user has not yet mastered the interest; the value is 1, indicating that the user has fully mastered the interest. The user will alas not change their interest in the process of diffusion. Alternatively,, in the process of event propagation, users will gradually forget the interest that has not been propagated, and at the same time, users will be interested in evolving their interest at the interval between the two diffusions.

Therefore, the level of interest at the end of event propagation at $t-1$ is different from that at the beginning of event propagation at t. In this paper, the interest satisfaction of users at t-time is modeled by using the matrix M_t^{Fv} ($d_v \times |I|$). Matrix M_t^{Fv} and $M_{t-1}{}^{v}$ have the same shape, M_t^{Fv} is made by $M_{t-1}{}^{v}$. After the interest evolution process, the t-time diffusion is defined as finished, and the system obtains the result of the user choice, according to the results of the diffusion, the DIEET model updates the interest master embedding matrix M_t^{Fv} to the $M_{t-1}{}^{v}$,at the end of the diffusion. This model is utilised in order to predict the performance regardingusers on the next candidate event e_{t+1}. Because there is a time interval between the end of t-time propagation and the beginning of $t+1$, the user's interest evolution behavior in the time interval will affect the interest satisfaction status. Therefore, $M_{t-1}{}^{v}$ should be updated to $M_{t+1}{}^{Fv}$ according to the factors affecting the evolution of interest, to predict the performance of users' diffusions.

Based on the aforementioned description, this paper defines the problem as: given the propagation history of each user, the following two goals are:

1) Achieved tracking changes in user interest status.
2) Forecast the performance of users on the next candidate event e_{t+1}.

IV. PROPOSED PREPROCESSING METHOD

This paper proposes a novel preprocessing method, which integrates user topic community discovery and influential spreaders identification, as shown in Figure 1.

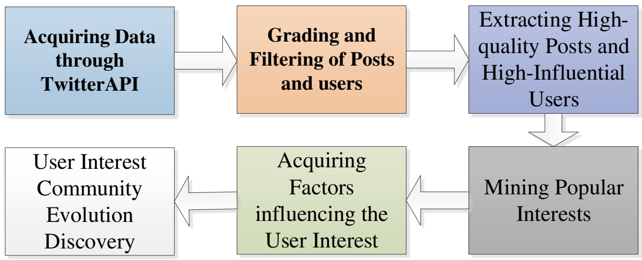


Fig. 1. Preprocessing method

A. Extracting Popular Interests and High-quality Users

Obviously, Hot interests will inevitably draw numerous high-quality users whose interests will frequently turn into popular ones [1]. Popular interests receive more high-quality user evaluations and recommendations than regular interests, and high-quality user recommendations can help spread this interest throughout social networks [2]. The HITS algorithm [12] has been described in this study, this takes into account the inextricable relationship between users and interests in order to extract popular interests and high-quality users. This will remove the users with fewer influence as well as unpopular interests. The authority score in the HITS algorithm represents the importance of the user, while the hub score represents the popularity of the interest. Influential spreaders are defined as users who have a high authority score. The final ratings for each person and interest may be determined iteratively and intuitively illustrate the relevance of the individual and the popularity of the interest.

B. Mining factors influencing the evolution of user interest

If users do not review what they have learned in time, their interest will evolve and their satisfaction of interest will decline, the retention rate of users' interests is affected by the following two aspects: the number of users' repeated transmission and the time interval. The time interval can be divided into the time interval of repeated transmission and the time interval of sequential transmission. In addition, according to the theory of memory trace decline in diffusion psychology [19], Therefore, this paper considers the following four factors that affect the evolution of interest.

- RT (repeated time interval): the time interval from the last spread of the same interest;
- ST (sequence time interval): the time interval from the last propagation;
- LT (repeated learning times): the number of repeated transmission of interests and hobbies;
- KL (past interest level): user's original satisfaction of the interest

RT, ST and LT are combined to get the vector $C_t\ (i) = [RT_t\ (i),\ ST_t\ (i),\ LT_t\ (i)]$, which represents the first three factors that affect the evolution of user's interest in interest *i*. The vector $C_t\ (i)$ corresponding to each interest constitutes the matrix $C_t\ (d_i \times |I|)$. This paper uses the vector $M_{t-1}{}^{v}\ (i)$ to measure the user's original satisfaction of interest *i*, the matrix $F_t = [M_{t-1}{}^{v},\ C_t]$ is obtained by combining the matrix C_t with $M_{t-1}{}^{v}$, which represents four factors that affect the evolution of interest.

C. User Interest Community Evolution Discovery

The communities within social networks evolve over time as a result of the interaction between the network's structure and the frequent interactions that take place within the structure.

Within this paper, referring to the subgraph increment method proposed by Jiang et al. [19], a novel User Interest Community Evolution, named UICE model based on cosine similarity is introduced in order to accurately detect the corresponding communities in the evolution of the user interest community. The steps included are as Fig. 2.

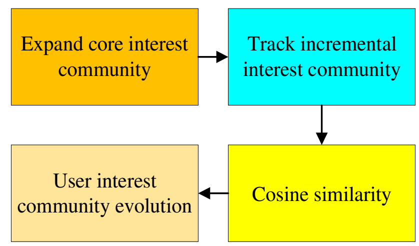


Fig. 2. The Procedure of the User Interest Community Evolution Model

V. TOPIC DIAGRAM CONSTRUCTION ALGORITHM

A. Word breaker and subject selection

The content within this document is organized into sections based on space and punctuation, and also stop words are removed after word segmentation. To minimize construction noise, conserve space, and balance the word frequency distribution, punctuation and also superfluous, modifiers are deleted. The relationship between news terms must be constructed using the topic as a link following word segmentation.

In this study, the TF-IDF (Term Frequency - Inverse Document Frequency) algorithm [1] is employed to rank the keywords of the news after word segmentation, and also the keyword with the highest weight is selected as the topic. The TF-IDF algorithm is a common weighting technique for information retrieval and text mining that evaluates the importance of a specific word in a piece of text statistically. TF represents the frequency of a specific word in a document, and the importance of a word is directly proportional to its frequency in the text (TF) while inversely proportional to its frequency in the corpus (IDF). The TF-IDF value is obtained by multiplying the TF value by the IDF value, and finally, the word with the highest TF-IDF value is selected as the topic.

B. Subject graph training

Natural language has specific relationships between words as they form sentences through various permutations and combinations to convey meaning. In order tobetter preserve the association information between words, this study utilizes the Neo4j graph database to store words after word segmentation.

A graph database is a type of NoSQL database that stores association information between things using graph theory, commonly seen in social network links between people. Graph databases have easy, quick, and efficient querying properties, where the nodes and connecting nodes' linkages are mainly stored. Within the corpus training process, we need to create an association table to record the relationship between different words, and to also save the topic information in the corpus into the word relationship in order to enhance the semantic information dimension related to the many-to-many combination relationships between words. The graph database meets the requirements for storing words and subjects mentioned above.

In this paper, WordGraph is built using a graph database to store words and also their relationships. Each word is constructed as a node, with the node's property becoming the word's text content. The text information of the topic is taken as the attribute of the connection, also every two related words are connected by a directional relation in the order of front and rear. The number of times the same relationship occurs is also saved as a relationship property.

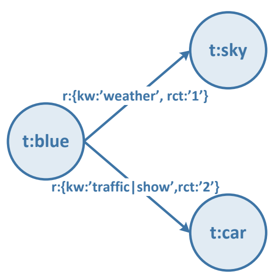


Fig.3. Word Relationships and Topics

The structure of word storage in graph database is shown in Figure 3, where "*t*" represents the word node and "*r*" represents the relationship between words 'Blue 'and' sky 'respectively represent two words that are connected in the same text. Both of their text topic' weather 'is stored in relation *r* as kW attribute, and their combination times in all texts are stored in relation *r* as *rct* attribute. Similarly, "blue" and "car" appeared twice in the two texts with the topics of "traffic" and "show".

Algorithm 1 Topic relation map training

Input: WordGraph [{(w_1)-[r:{kw:'kw_1',rct:'1'}]->(w_2)},…], a set of words and relations, Eventslist [{w_{11},w_{12},w_{13},…w_{1n}},…], a set of events.

Output: Topic relation map

```
foreach News ∈ Eventslist do
  kw_j=News.TF-IDF()
end
foreach w_i,w_i+1 ∈ News do
 if !WordGraph.contains(w_i,w_i+1) do
   WordGraph.add(w_i,w_i+1)
  r = WordGraph.match({(w_i)-[r]->(w_i+1)}))
  if r is null do
   WordGraph.add(r:{kw:kw_j,rct:'1'})
  else do
    r.rct++
    if(!r.kw.contains(kw_i))
     r.kw+= kw_i
  end
 end
end
```

In each computation cycle of the eventslist's cumulative training procedure, two neighboring words are analyzed. The relevant word is not added to the graph if the current word already exists in WordGraph. The directional relationship between words is also defined by the location of the words in the text. If a relationship in the same direction already exists in WordGraph, the relationship will not be added in the follow-up training, but the *rct* attribute in the relationship is added by 1, and *rct* represents the number of times the word pairs in the same direction appear. If the subject of the two words is not included in the *kw* attribute, it will be added additionally. For example, if the original subject is *kw*1 and the current subject is *kw*2, the representation of the attribute *R* is: *R*: *r*:{*kw*:'*kw*1|*kw*2',*rct*:2}.

Sentences with a high degree of resemblance frequently have comparable, although not always identical, themes. As a result, it's critical to include subject similarity when determining sentence similarity. Because the TD-IDF algorithm selects the subject from the word set, the topic words are also kept as nodes in the WordGraph graph database. The best path (BP) connecting subjects in WordGraph is the key to judging topic significance.

Because two words with low correlation will traverse practically the whole WordGraph in BP search, a threshold value is required to enhance system performance on the number of relationship layers. The 'Six-degree space theory' is introduced within this study in order to give a theoretical foundation for threshold setting.

The six-degree space theory, also known as the six-degree division theory or the small world theory, is a mathematical hypothesis. According to the hypothesis, in social networks, there are no more than 6 persons between a person and any stranger, implying that each stranger can be met through a maximum of 6 intermediates. This is because WordGraph has a structure comparable to social networks, any two different words may always be related by jumping along a path within six levels.

VI. DIEET MODEL

This paper proposes a deep event tracking model DIEET, which integrates propagation and interest evolution, as shown in Figure 4.

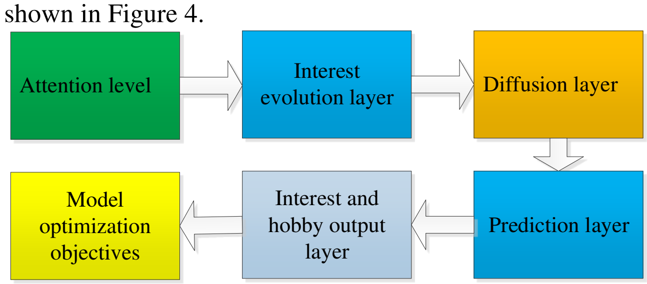


Fig. 4. DIEET model

Our DIEET model is divided into a attention layer, forget layer, prediction layer and also learning layer. This paper proposes a deep event tracking model DIEET, which integrates propagation and interest evolution, as shown in Figure 4.

Also the interest level output layer: known as the attention layer takes the set d_i of interest and interests covered by the events and used as input to calculate the weight of the interest related to the events and interests; the interest evolution layer calculates the interest related weights of the events and the interests; the interest evolution layer will utilise the four factors that influence the evolution of the interest proposed in this section in order to make the user last M_{t-1}^v. Ultimately at the end of time diffusion, the interest master embedding matrix M_{t-1}^v is updated to the interest master embedding matrix M_t^{Fv} at the beginning of the current time diffusion M_t^v, effectively leaving the user. Also Forgetting behavior is modeled; the prediction layer predicts the performance of users' choices according to M_{t+1}^{Fv} when $t+1$ time starts to choose questions; the diffusion layer will utilise LSTM network in order to predict the performance of the users' choices. The M_t^{Fv} at the beginning of diffusion is updated to M_t^v at the end of this propagation, which then models the user diffusion behavior; the interest output layer uses the user to also spread this time. Ultimately at the end of the paper, M_{t-1}^v is used as input, and the user interest vector $value_t$ is designated as output, and also the user's satisfaction of each interest is output in real time at the same time.

1) Attention level

The function of attention layer is to calculate the interest related weight between events and interests. The input of the attention layer is classed as the user's current event topic e_t and the interest set I_t involved in the topic. In order to map e_t to the continuous vector space, this paper uses e_t times event embedding matrix $A(d_i \times |E|)$ to generate a d_i dimension event embedding vector v_t, each d_i vector in matrix A is the embedded representation of an event. Experts mark the events and interests covered by each event and store them in the set I_t. This paper filters out the unrelated interests through the K-filter, and retains the events covering interests and interests. Matrix I_t stores the embedded vectors of interest and interests covered by the events, amongst which, each d_i dimension vector in matrix I_t is an embedded vector of interest and interest covered by events. The internal product between event embedding vector e_t and interest inclusion embedded vector I_t (i) is calculated, and then the softmax value of the internal product is calculated and stored in vector RS_t. Vector RS_t. represents the interest related weight between event e_t and interest and interest covered by event, as shown in formula (1):

$$RS_t(i) = Softmax\left(e^{'}_t R_t\left(i\right)\right) \tag{1}$$

where the $|I|$ dimension vector w_t is the interest related weight vector between event and all interests. Since the problem cover interests and hobbies are filtered out by the interest filter, the model needs to implement the interest related weight of the events covering interests and hobbies into the corresponding position of w_t. first, initialize the $|I|$ dimension all zero vector w_t, namely $w_{t=}$ [0, 0]; then, implement the weight of each dimension of RS_t into the corresponding position of w_t, namely w_t $[I_t$ [i] $= RS_t$[i], and get the weight of events and interest related to each interest.

2) Interest evolution layer

The interest evolution layer, based on the factors influencing the evolution of interest described in previous, F_t, embeds the embedded matrix M_{t-1}^v of interest satisfaction of users at the end of the last diffusion

The interest evolution process is carried out in order to obtain the interest satisfaction embedding matrix M_t^{Fv} at the beginning of this diffusion. Inspired by the interest evolution gate and input gate in LSTM, the image is then obtained. When the factor of interest evolution of interest is updated by F_t, the original information in M_{t-1}^v must be erased first before being written. Regarding the modeling of interest evolution behavior, erase is determined as the decline of the degree of interest and interest satisfaction of process control users, and also the updating of the degree of interest and interest satisfaction of process control users.

Utilising a full connection layer with sigmoid activation function, the factor $F_t(i)$ of influencing the evolution of interest in interest i of users is transformed into interest hobby i corresponding memory erase vector $\boldsymbol{fe}_t(i)$:

$$fe_t(i) = Sigmoid(FE_T\mathbf{F}_t(i) + b_{fe}) \tag{2}$$

where the shape of the weight matrix FE of the whole connection layer is $(d_v + d_c) \times d_v$, and the offset vector BFE of the whole connection layer is d_v dimension. The memory erase vector $\boldsymbol{fe}_t(i)$ is classed as a dimension. Regarding the column vector with degree DV, all the values in the vector are within the range of (0,1). Utilising a full connection layer with tanh activation function, the factor $F_t(i)$ which affects the user's interest evolution degree will ultimately be transformed into the memory update vector $\boldsymbol{fa}_t(i)$ corresponding to the interest i:

$$fa_t(i) = Tanh(FA_T\mathbf{F}_t(i) + b_{fa}) \tag{3}$$

where the shape of the weight matrix FA of the whole connection layer is $(d_v + d_c) \times d_v$, the offset vector b_{fa} of the whole connection layer is d_v dimension, and the memory update vector $\boldsymbol{fa}_t(i)$ is a dimension M_t^{Fv}. The column vector with degree d_v is obtained from M_{t-1}^v update according to memory erase vector and memory update vector M_t^{Fv}.

$$M_t^{Fv}\left(i\right) = M_{t-1}^{v}\left(i\right)\left(1 - fet\ (i)\right)\left(1 + fat\ \left(i\right)\right) \tag{4}$$

The evolution layer outputs the interest master embedded matrix M_t^{Fv} at the beginning of this diffusion. The interest evolution layer can also model the evolution process of users' interest in various interests and hobbies respectively. According to the different history of users' interest and interest's diffusion, the degree of interest evolution of users for different interests and hobbies is calculated.

3) Diffusion layer

The diffusion layer tracks the interest satisfaction changes in the process of diffusion according to the results of the user choices. The interest master embedding matrix M_t^{Fv} is updated to the embedded matrix M_t^v at the end of the user diffusion, and the propagation behavior is modeled. Tuple (ET, RT) represents the choice result of the user in time t. in order to map tuples (e_t, i_t) to continuous vector space, tuples (e_t, i_t) are multiplied by the result of the choice with $d_v \times 2|E|$ embedded in matrix B, and the d_v dimension choice results are embedded in Vector s_t. the diffusion layer also embeds the choice result into the interest related weight vector w_t corresponding to the vector s_t and also the event as input, and the utilises the user within the online education system The results resonating from this practice are used as indirect feedback of the user diffusion effect. The interest satisfaction status in the process of user diffusion is updated through LSTM network, and the diffusion behavior is modeled.

$$M_t^v(i) = LSTM\ (s_t, w_t(i) M_t^{Fv}(i); \theta) \tag{5}$$

4) Prediction layer

The purpose of the prediction layer is to predict the user performance on the next candidate event e_{t+1}. Because the user's interest evolution behavior within the interval of two choices will affect the interest satisfaction state, the prediction layer predicts the probability of users to correctly choice the event e_{t+1} according to the interest of the choice matrix M_{t+1}^{Fv} when the current time starts to choose the question. The weighted sum of interest related weight w_{t+1} and the user interest satisfaction at the beginning of the current time is embedded into the weighted solution of M_{t+1}^{Fv}. The vector d_{t+1} is the weighted satisfaction degree of the user for the interest and interest of the problem.

$$d_{t+1} = \Sigma i w_{t+1}(i) M_{t+1}^{Fv}(i) \tag{6}$$

The successful solution of the problem by experienced by users is not only related to the comprehensive satisfaction of the user's interest and hobbies, but also the event itself. Therefore, the combination vectors $[d_{t+1}, v_{t+1}]$ connected by vector d_{t+1} and v_{t+1} are input into the full connection layer with the *Tanh* activation function, Output vector h_{t+1} includes the comprehensive satisfaction of the user's interest and interest in the events and the characteristics of the events themselves. Amongst them, matrix W_1 and vector b_1 represent the weight and offset of the full connection layer respectively.

$$h_{t+1} = Tanh\left(W_1^T \left[d_{t+1}, v_{t+1}\right] + b_1\right) \tag{7}$$

Finally, the vector h_{t+1} is input into a full connection layer with sigmoid activation function, and the probability P_{t+1} which indicates the user choices also indicates the problem correctly alleviated.,

Matrix W_2 and vector B_2 represent the weight and offset of the full connection layer respectively:

$$P_{t+1} = Sigmoid\ \left(W_2^t h_{t+1} + b_2\right) \tag{8}$$

5) Interest output layer

The output layer of interest takes the embedded matrix M_t^v of interest satisfaction at the end of user diffusion and designated as an input. The output represents the $|I|$ dimension vector value of the interest satisfaction level at the end of the diffusion. Each dimension of the vector is between (0,1), indicating the user's satisfaction of the interest.

In the prediction layer, each time node t, formula (7) and formula (8) predict the performance of users on a specific event set according to dual inputs: users have obtained comprehensive interest of the embedding vector d_t and also the event embedding vector v_t for the event cover interests and hobbies. Therefore, if you just want to estimate the input of no event questions. The user can omit the events embedded into v_t, and let the user know the M_t^v of the I column of the embedded matrix M_t^v (i). As an input of the equation, specifically, after the output matrix M_t^v of the diffusion layer, in order to estimate the satisfaction of the I interest, the weight vector △i= (0, …, 1) is constructed, where the value of dimension I is equal to 1, and the embedding vector M_t^v (i) of the first interest is extracted by formula (9), and then the interest satisfaction level is estimated by formula (10) and formula (11):

$$M_t^v(i) = \beta_i M_t^v \tag{9}$$

$$y_t(i) = Tanh\left(W_1^T \left[M_t^v(i), o\right] + b_1\right) \tag{10}$$

$$value_t(i) = Sigmoid\ \left(W_2^t y_t(i) + b_2\right) \tag{11}$$

where vector o= (0, 0, …, 0) which is the same as the event embedded v_t dimension, this is used to supplement the vector dimension. Parameters W_1, W_2, b_1, b_2 are identical with formula (7) and formula (8). Calculated is the satisfaction degree of each interest in the interest space in sequence, and ultimately receive the user interest satisfaction vector value

6) Model optimization objectives

The parameters to be trained in the model are event embedding matrix A, choice result embedding matrix B, neural network weight with bias and interest matrix M^I. This paper optimizes the parameters by ultimately minimizing the cross-entropy loss function between the predicted value p_t of the model for the user choice result and also the real result i_t of the user choice. The loss function is shown in formula (12). Adam is used in this paper Methods the parameters were optimized.

$$L = -\sum_t \left(r_t log\ p_t + (1 - r_t) log\left(1 - p_t\right)\right) \tag{12}$$

VII. EXPERIMENTS

A. Experiment Datasets

We generated our dataset from Twitter (http://twitter.com/) via the Twitter API. This dataset consists of 60,000 posts from October 21–28, 2015. As discussed above, in order to reduce the impact of the bump phenomenon, we included that only those users who have been published or have commented upon posts in our dataset.

There areapproximately 6 events detected through our event detection algorithm.
Regarding data processing, we extract keywords and summaries from events through an improved HITS algorithm [11]. In order to demonstrate our method in a improved manner, we choose key posts as a topic abstract of each event, within this paper, figure 1 shows the details regarding these events.

B. Comparative Methods

DIEET, is identified as deep event tracking model that integrates propagation and interest development, this model is proposed within this study. At the same time, the model overcomes the problem of the ability to replace influential spreaders whenever users' interests change, which is common in the evolution of interest communities. The proposed method's performance is assessed by current algorithms comparison based on the aforementioned factors.

First aspect: dynamic user interest community evolution tracking:

(1) FacetNet algorithm [13]: This algorithm produces associations utilising the random block model and also analyzes the evolution of associations using the Dirichlet probability model.

(2) DSBM algorithm [16]: This method is a Bayesian inference-based approach used for detecting communities in dynamic social networks and also recording community dynamics.

(3) CIDLPA algorithm [12]: This algorithm divides nodes into two categories according to the degree of influence in order to detect communities in dynamic social networks.

Second aspect: Core user replacement based on subgraph matching:

(1) NeMa algorithm [22]: The designated algorithm is a unique subgraph which match's graph query architecture that evidently allows for ambiguity in both structure and node labels.

(2) DYNMOGA algorithm [15]: An algorithm that automatically discovers the number of communities and also selects the ultimate answer using the evolutionary method.

(3) SM-RDF algorithm [23]: This approach is the same as searching for subgraphs of fuzzy graphs that contain a greater opportunity of matching a query graph.

(4) DyPerm algorithm [47]: This algorithm is the first dynamic community detection method which ultimately optimizes a novel community scoring metric.

(5) iDBLINK algorithm [22]: This algorithm has the ability to update the local link community structure in the current moment through the change of similarity between the edges at the adjacent moments, this includes the creation, growth, merging, deletion, contraction, and division of link communities.

(6) EEUICD algorithm [48]: This approach employs a genetic algorithm which is a nature-inspired algorithm in order to improve the quality of community discovery.

C. Evaluation Measures

In order to compare the proposed DIEET model with various other methods, validation measures are introduced: This is designated as the normalized mutual information (NMI):

$$NMI(A,B) = \frac{-2\sum_{t=1}^{C_A}\sum_{j=1}^{C_B} C_{ij}log(C_{ij}N/C_{i.}C_{.j})}{\sum_{t=1}^{C_A} C_{i.}log(C_{i.}/N) + \sum_{t=1}^{C_B} C_{.j}log(C_{.j}/N)} \tag{13}$$

where C_A denotes the number of communities in A, and C_B denotes the number of communities in B, $C_{i.}$ denotes the total number of rows in matrix C, $C_{.j}$ denotes the total number of columns in matrix C, and also N denotes the number of nodes.

In order to compare the proposed DIEET model with other methods, three validation measures are introduced: community precision (*Precision*), community recall (*Recall*), and also *F-measure*:

$$Precision = \frac{|x \cap y|}{|x|} \tag{14}$$

$$Recall = \frac{|x \cap y|}{|y|} \tag{15}$$

$$F-measure = 2\frac{Precision*Recall}{Precision + Recall} \tag{16}$$

The area under the curve AUC is designated as an important index to measure the quality of community evolutionary algorithm. AUC compares the probabilities of existing links and unlinked links in the test set and records them as pre_{linked} and also $pre_{unlinked}$. If $pre_{linked} > pre_{unlinked}$, then the evaluation index accumulates 1 point; If $pre_{linked} = pre_{unlinked}$, the cumulative score of the evaluation index is 0.5; AUC is calculated as shown in formula (18).

$$AUC = \frac{(n' + 0.5n'')}{n} \tag{17}$$

where n is the number of measurements and n' is $pre_{linked} > pre_{unlinked}$ value, n'' is $pre_{linked} = pre_{unlinked}$ value. The higher the AUC value, the better the quality of the link prediction algorithm.

D. Experiment Setting and Results Analysis

We conducted our experiments with a computer complete with an Intel I7 3.4 GHz CPU and 32 G memory.

We tuned the parameters via a grid search. For LDA, α = 0.5 and β = 0.1. In all the experiments, we used Gibbs sampling for 1,000 iterations. The results reported comprised of an average of ten runs. In the process of filtering high-quality posts, we set all of the initial authority scores *d.a* and hub scores *u.h* to 1.

The number of posts (K). Due to the irregular writing of short text on social networks. Th posts consist of a large amount of noise in the event tweet stream. That is to say, the event tweet stream textual content does not always relate to an event. Therefore, we first try to extract a small number of high-quality posts for the event. Before constructing the sub-event in our experiment, we also need to extract posts from the subevent tweet stream. We run topic decision graph [33] to automatically obtain key posts from the sub-event tweet stream and then extract top-K determined in theupper right region as the sub-event's posts set.

TABLE I. RESULTS FOR EVENT DETECTION

Real-life Event	Key post	Created Time
The rise and controversy of classical economics	@namasteacup "classical economics"is specifically in the text:p	Tue Oct 27 16:56:20 2015
Economic deficit in United States	@Shamsher1111 @johnefrancis If you have a master in economics and don't understand uses of deficit spending, you are a very great fool.	Tue Oct 27 19:49:23 2015
Economic crisis in Poland	@BeingAnkit_ My mind starts boggling at Economics. I better leave you to study.	Tue Oct 27 07:08:20 2015
The rise of cultural economics	"each part has a size measuring its efficiency economics became more efficient than culture for organizing society @enleuk"	Mon Oct 26 23:25:51 2015
The rise of cultural economics	@NYSELaxative @StartlinglyOkay What kind of input, output and filter? And economics is limited to property and only one part of culture.	Mon Oct 26 21:03:19 2015
The rise of football economics	@ArsenalReport @JanuzajA11 @Firzaapras err I study economics so I I'd know about this subject especially, and its a fact that it doesn't	Mon Oct 26 16:07:03 2015
The rise of football economics	@mk_9873 @januzaja11 @firzaapras Maybe because you only hang out with mouth breathers? It's how all economics work, not just football.	Mon Oct 26 16:05:42 2015
Economic crisis in Ireland	@UB_Economics I am sorry to hear this @hazeyhall, have you managed to arrange an appointment now? PH	Mon Oct 26 11:15:12 2015

TABLE II. PREDICTION RESULTS OF MODELS

Model	AUC	
FacetNet	0.596	0.698
CIDLPA	0.714	0.778
DSBM	0.745	0.780
DIEET	**0.749**	**0.801**
DIEET(without forget layer)	0.731	0.788

E. Result Analysis

(1) Analysis of event tracking results

Table II displays the average AUC, DIEET, and also indicate three comparison techniques findings. Identified within the twitter datasets, the prediction performance of DSBM is determined as the lowest, recognising that the DSBM has limitations in modeling users' potential interests determined to be binary variables. The average AUC of DIEET is identified as greater than the other three comparison methods, this indicates that the proposed DIEET model is designated as superior to the existing models when predicting users' future performance. FacetNet models users' overall interests using the potential vector of a recurrent neural network, alas this cannot model users' happiness with each interest. Alternatively, FacetNet's prediction performance is determined as poorer than that of CIDLPA, DSBM, also the DIEET. CIDLPA, DSBM, and DIEET may all model the user's level of satisfaction with each interest point, alas the CIDLPA and DSBM chooses to ignore the user's interest evolution behavior during the dissemination period, and also the default user's satisfaction degree of unrelieved interests is always the same**.**

Additionally, this research evaluates the DIEET model's prediction performance after the deletion of the interest evolution layer. Interest will not vary in this case when it is deemed the forget behavior as the default user in the diffusion interval. The model's prediction performance has also been deemed poorer when the interest evolution layer is removed. This demonstrates how the user's interest evolution behavior ultimately affects their interests. The designated elements that influence interest evolution suggested in this research are also quite effective.

It is common knowledge that, the basic task of event tracking is designated as the output of the user's satisfaction of various interests in real time. Ultimately, this paper conducts the following experiments to verify the effectiveness of the DIEET model within the event tracking task**.**

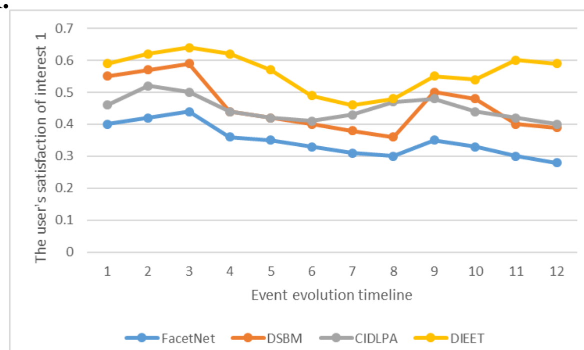


Fig. 5. The tracking of user's satisfaction degree of interest 1

This paper identifies intercepts within the propagation records of a user within the twitter dataset over a period of time, and also uses the DIEET model and also the existing three models in order to track the changes of the user's satisfaction degree of interest 1, as shown in Figure 5. The horizontal axis within the figure represents the user's propagation history, and also the k_t in the meta group (k_t, r_t) represents the user's propagation interests, RT also represents the user's response. The vertical axis represents the satisfaction degree of interest 1 in relation to model tracking**.**

The experimental results indicate that: after the user selects interest 1 correctly, the tracking results of DIEET and also the existing three models on the satisfaction degree of user's interest 1 improve (the output value increases); after the user selects the interest 1 correctly, also the tracking results of all models on the satisfaction degree of user's interest 1 improve (all models have lower tracking results on the satisfaction of user interest 1); and also after the user selects the interest 1 correctly, the tracking results of all models on the satisfaction degree of user's interest 1 improve The foregoing findings indicate that the existing three models, as well as the DIEET model, have been seen to gain traction. The findings reveal that all models have the ability to model the user's diffusion behavior based on the user's decision outcomes for the third time.

Illustrated in Figure 5, a selected user has still not reviewed the interest 1 after an 8^{th} selection, and the corresponding interest satisfaction degree value of FacetNet and CIDLPA have a minute change after the 8^{th} time, with the results showing that: Our proposed DIEET model can model the user's interest evolution behavior and also track the degree of change relating to the user's interest satisfaction degree value caused by forgetting behavior; Alas FacetNet, CIDLPA and DSBM all ignore the user's interest evolution behavior and therefore cannot track the user's interest changes caused by interest evolution.

Users will have the ability yo update the true satisfaction degree of the related interests which are based on the diffusion outcomes, in order to represent the user's diffusion process. The other three models do not have the ability to update the user's choice of interest 1, whereas the DIEET model shows that the user's satisfaction with interest 1 has been seen to be declining. The above results illustrate that both the DIEET model and the existing three models can model user diffusion behavior, but also the existing three models cannot model user interest evolution behavior, whereas the DIEET model has the ability to model user interest evolution behavior and also track user's choice of each interest in real time.

Within this paper, the propagation records of three users in a period of time are randomly selected from the test data set, and also the DIEET model outputs their satisfaction of five hobbies. The partial choice sequences of the three users are as follows:

Ha=[(1,1),(2,1),(3,1),(4,0),(5,0),(6,0),(1,0),(2,1),(3,1),(4,0),(5,1),(6,1)];
Hb=[(6,1),(5,1),(4,1),(3,1),(2,1),(1,1),(2,1),(3,1),(4,1),(5,1),(6,1),(1,1)];
Hc=[(1,1),(2,1),(3,1),(4,0),(5,1),(6,1),(1,1),(2,1),(3,0),(4,1),(5,0),(6,0)].

Each item (KT, RT) in the sequence represents the result of the user's choice, in which KT represents the interests and hobbies covered by the user's current event, RT represents the result of the user's choice, 0 represents the incorrect choice, and 1 represents the correct choice.

(2) The event evolution analysis

As is shown in Table III and IV, we can easily identifythe hot event evolution chains during event evolution and also the influential spreaders' interests changing. The reason for this is the DIEET model is proposed to judge correlation between events, which have the ability to update the seed users according to their changeable interest in time, which can greatly improve the precision of event tracking. simultaneously, we can also discover which hot events comprise of long interests evolution chain, which ultimately indicate the popularity of the hot events compared with the HEE model [49].

TABLE III. THE RESULTS OF KEY POSTS DETECTION BASED ON DIEET MODEL

Time	*Key Post*	*Event evolution relationship*
Mon Oct 26 2015	P1	
	P2	E1→E2
	P3	E1→E2
	P4	E1→E2→E3
	P5	E1→E2→E3
Tue Oct 27 2015	P6	E3→E4
	P7	E3→E5
	P8	E3→E4→E5→E6

TABLE IV. THE RESULTS OF INFLUENTIAL SPREADER'S INTERESTS DETECTION BASED ON DIEET MODEL

Time	*Hot Event*	*Interest evolution relationship*
Mon Oct 26 2015	E1	
	E2	I1→I2
	E2	I1→I2
	E3	I1→I2→I3
	E3	I1→I2→I3
Tue Oct 27 2015	E4	I1→I2→I3 →I4
	E5	I1→I2→I3 →I5
	E6	I1→I2→I3 →I6

(3) Case study

This section presents a case study of the 'economics event' in order to explain our design's effectiveness further.

As shown in Table III, we can easily discover the events evolution chains, 'E3→E4→E5→E6', which are based on the DIEET model. The reason is because a topic decision-based event evolution chains detection method is proposed to judge correlation between key posts (marked as P1 to P8), which can detect the process of event evolution. Simultaneously, we can also discover which hot events have a long event evolution chain, which ultimately indicate the popularity of hot events compared with the-state-of-the art models [12,13,15,16].

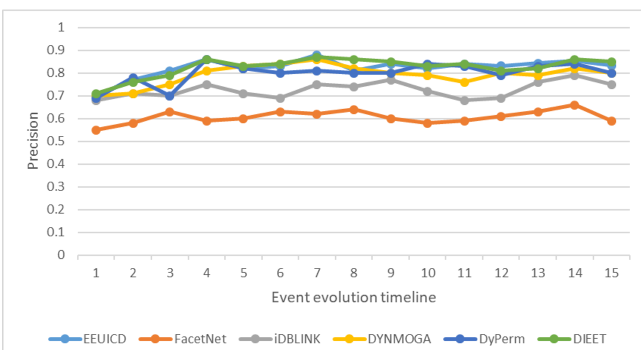
Furthermore, it is illustrated within Table IV, we can discover that the key users' interests (marked as I1 to I6) will be modified over time, which then validates our proposed DIEET model in relation to detecting greater interests for each user within users' community, ultimately solving the big problem of data sparsity within microblogging networks. Simultaneously, within our proposed automatic topic clustering algorithm, it has been identified that all short texts can be combined into clusters with similar topics. also with the improved user-interest model all short texts in each cluster can be integrated to form a long text document simplifying the determination of the overall topic in relation to the interest distribution of each user during the evolution of hot events. This addresses the problem of sparse data and improves the quality of topic definition and the accuracy of user interest discovering.

(4) Precision rate, Recall rate and F-measure value comparison

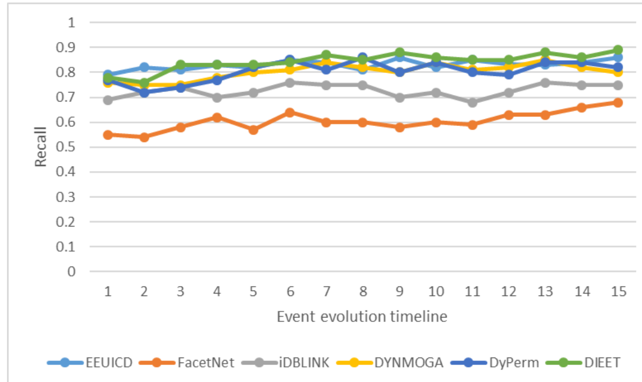
We compare the DIEET model with other existing algorithms, with the selection Precision rate, Recall rate and F-measure value, which are illustrated above as the evaluation criteria. The result is illustrated in Figure 6.

The accuracy of the FacetNet method has been significantly lower than the other four algorithms, as seen in Figures 6(a), 6(b), and 6(c). This can potentially be because the FaceNet algorithm generates associations using a random technique and also examines their evolution using the Dirichlet distribution method, which results in relatively poor accuracy. Although the IDBLink method somewhat compensates for the inadequacies of the FaceNet approach, there is a requirement for a procedure for initialization and optimization based on community discovery.

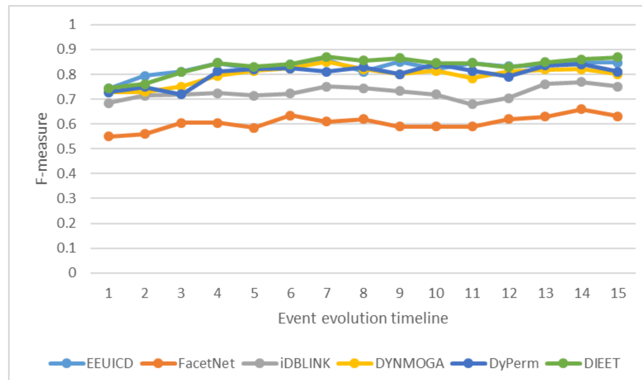
The DyPerm method corrects the flaws in the iDBLink algorithm by maximizing persistence to achieve the network community structure and also locally adjusting incremental nodes to discover the optimal community ownership. Alternatively, the DyPerm approach, is time expensive and also takes a long time to calculate the persistence index. Without taking efficiency into account, the DyPerm algorithm has shown to be able to produce good results. The DYNMOGA algorithm corrects the flaws of the preceding technique by employing a genetic algorithm in order to determine the number of communities automatically. It has produced positive results and also has a considerably greater efficiency than the DyPerm algorithm. The EEUICD algorithm utilizes a subgraph matching-based community recognition and an evolution algorithm, also it consists of a genetic reinforcement mechanism to match community dynamic similarities. The EEUICD method outperforms other algorithms in terms of optimization outcomes.



(a)



(b)



(c)

Fig. 6. Precision rate, Recall rate and F-measure value comparison

Figure 6 illustrates that the DIEET model's interest community detection is deemed more accurate than other comparison techniques. The reason is potentially because the improved algorithm based on label propagation is used in the process of user interest community discovery to improve the overall clustering effect and also convergence speed, also the preprocessing model based on influence and similarity is utilised to effectively improve the local information quality of users and posts.

(5)Tracking effect analysis of dynamic user interest community evolution

The user data within the Twitter dataset not only reflects each individual user, but also the forwarding, reply, and other link relationships that exists between them. The DIEET method is compared to various current algorithm models, employing the NMI index as the assessment benchmark. Figure 7 depicts the experimental comparison findings.

In Figure 7, the NMI value of DIEET is occasionally lower than that of EEUICD and DYNMOGA in the beginning of the study, but seen at time point 2, the NMI value of DIEET is consistently greater than that of EEUICD, FaceNet, DyPerm, and iDBLink. The NMI value of EEUICD, on the other hand, is occasionally higher than that of DIEET. The reason is potentially because the EEUICD label propagation algorithm uses a cascade information diffusion model, which potentially increases the accuracy of the label propagation technique and also mitigates the negative effects of the random block model.

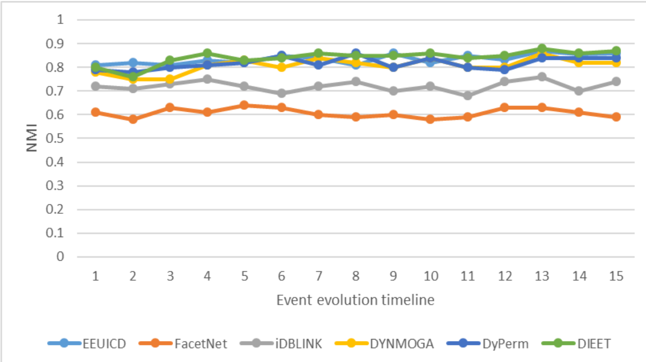


Fig.7. NMI comparison

We can see from the experimental findings that the FacetNet method has been shown have the lowest feedback due to the fact it poorest because generates associations using a random block model and also analyzes the evolution of associations using a probability model based on the Dirichlet distribution, resulting in low accuracy. As a result, the FacetNet algorithm is deemed unsuitable for discovering user interest communities within dynamic social networks. Despite the fact that the iDBLink algorithm solves the limitations of the FacetNet algorithm, there is still a problem with lack of an initialization technique and also an optimization strategy, resulting in an unsatisfying output. Despite demonstrating low efficiency, the DyPerm method has produced good results by increasing persistence in order to create the network community structure and also optimizing incremental nodes locally. The EEUICD algorithm adopts the community evolution recognition algorithm which is based on subgraph matching. Although it effectively identifies the user's interest community, it sometimes has the problem of low accuracy of community recognition which is due to the lack of prediction of user interest behavior.

The DIEET model, on the other hand, has shown to correct the flaws of previous algorithms. To successfully increase the quality of local information, the HITS algorithm is applied to the preprocessing model. The clustering effect and convergence speed are also increased utilizing an improved label propagation technique, and also an adequate tracking effect of interest community evolution is realized. As a result, the DIEET algorithm not only maintains the stability of the NMI value recognized by user interest communities, but it also improves its scalability, reduces the execution time of user interest community discovery, and also resolves the trade-off between the number of communities and the two objective functions. As a result, the DIEET model has shown to outperform other comparison models. The experimental findings also reveal that the DIEET model has shown the greatest user interest community discovery performance.

VIII. CONCLUSION

The influence of users' interest evolution behavior on their interest fulfillment is the designated topic of research demonstrated in this paper, this introduces the DIEET model, which ultimately integrates propagation and interest evolution. It also outperforms typical event tracking algorithms in forecasting user decision performance, according to our investigations. When tracking changes in users' interests and hobbies, it has shown to not only track the change of user's interest in the diffusion process based on the user's choices, but also reflect the user's interest evolution behavior caused by complex interest evolution factors, and also track the change process of user's interest caused by the user's interest evolution in real time.

In the future research, we will explore the following aspects.

(1) There has shown to be prior, posteriori and inclusion relationships among interests, so there is a need to elucidate the relationship characteristics between interest points, and then accurately infer the user's satisfaction of each interest;

(2) The data set analyzed within this paper involves relatively few interests and hobbies, so it has been found to be simpler to track the user's satisfaction of the events according to the user's choice results. For the events with a high number of interests and hobbies, such as comprehensive events, it will potentially lead to uncertainty of the satisfaction of interests and hobbies, this is potentially our next research project.

DECLARATIONS

Ethics Approval

Written informed consent for publication of this paper was obtained from the Suqian University, Jiangsu University , University of Leicester and all authors

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, author- ship, and/or publication of this article.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contribution

Jun Ge: Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. Lei-lei Shi: Conceptualization, Investigation, Data curation. Lu Liu: Funding acquisition, Validation, Writing – review & editing. Zi-xuan Han: Validation, Writing – review & editing. Anthony Miller: Data curation, Validation, Funding acquisition, Writing –review & editing.

Funding

The work reported in this paper has been supported by the China Postdoctoral Science Foundation (Grant number: 2023M731368), the Natural Science Foundation of the Jiangsu Higher Education Institutions (Grant number: 22KJB520016) and Jiangsu University Innovative Research Project (KYCX22_3671).

Consent to publish

Written informed consent was obtained from all authors for publication of this study and any accompanying images.

All authors of this article gave their consent for contribution. Copies of this written consent are available for inspection by the Editor-in-Chief of the Journal.

REFERENCES

[1] L.-l. Shi, L. Liu, L. Jiang, R. Zhu, and J. J. N. Panneerselvam, "QoS prediction for smart service management and recommendation based on the location of mobile users," Neurocomputing, vol. 471, pp. 12-20, 2022.

[2] D. Mumin, L.-L. Shi, L. Liu, and J. Panneerselvam, "Data-driven diffusion recommendation in online social networks for the internet of people," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, pp. 166-178, 2022.

[3] J. Tong, L. Shi, L. Liu, J. Panneerselvam, and Z. Han, "A novel influence maximization algorithm for a competitive environment based on social media data analytics," Big Data Mining and Analytics, vol. 5, pp. 130-139, 2022.

[4] L.-L. Shi et al., "Human-Centric Cyber Social Computing Model for Hot-Event Detection and Propagation," IEEE Transactions on Computational Social Systems, vol. 6, no. 5, pp. 1042 - 1050, 2019.

[5] K. R. Žalik, "Evolution algorithm for community detection in social networks using node centrality," in Intelligent Methods and Big Data in Industrial Applications: Springer, 2019, pp. 73-87.

[6] M. Sattari and K. Zamanifar, "A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks," Journal of Computational Science, vol. 25, pp. 122-133, 2018.

[7] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "QoS Prediction for Service Recommendation with Deep Feature Learning in Edge Computing Environment," Mobile Networks & Applications, vol. 25, pp. 391-401, 2020.

[8] M. Li, Q. Lu, and M. Zhang, "A Two-Tier Service Filtering Model for Web Service QoS Prediction," IEEE Access, vol. 8, pp. 221278-221287, 2020.

[9] R. Nagarajan and R. Thirunavukarasu, "A Service Context-Aware QoS Prediction and Recommendation of Cloud Infrastructure Services," Arabian Journal Forence & Engineering, vol. 45, pp. 2929–2943, 2020.

[10] J. Li and J. Lin, "A probability distribution detection based hybrid ensemble QoS prediction approach," Information Sciences, vol. 519, pp. 289-305, 2020.

[11] L. Shi, Y. Wu, L. Liu, X. Sun, and L. Jiang, "Event detection and identification of influential spreaders in social media data streams," Big Data Mining & Analytics, vol. 1, pp. 34-46, 2018.

[12] J. Ge, L. L. Shi, L. Liu, and X. Sun, "User Topic Preferences based Influence Maximization in Overlapped Networks," IEEE Access, vol. 8, pp. 141328-141345, 2019.

[13] S. Tang, S. Yuan, and Y. Zhu, "Convolutional Neural Network in Intelligent Fault Diagnosis Toward Rotatory Machinery," IEEE Access, vol. 8, pp. 86510-86519, 2020.

[14] S. Tang, S. Yuan, and Y. Zhu, "Deep Learning-Based Intelligent Fault Diagnosis Methods Toward Rotating Machinery," IEEE Access, vol. 8, pp. 9335-9346, 2020.

[15] S. Zeng, B. Zhang, Y. Lan, and J. Gou, "Robust collaborative representation-based classification via regularization of truncated total least squares," Neural Computing & Applications, vol. 31, pp. 5689–5697, 2019.

[16] B. Yuan, J. Panneerselvam, L. Liu, N. Antonopoulos and Y. Lu, "An Inductive Content-Augmented Network Embedding Model for Edge Artificial Intelligence," in IEEE Transactions on Industrial Informatics, vol. 15, no. 7, pp. 4295-4305, 2019.

[17] M. Nilashi, P. F. Rupani, M. M. Rupani, H. Kamyab, W. Shao, H. Ahmadi, et al., "Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach," Journal of Cleaner Production, vol. 240, p. 118-162, 2019.

[18] Z. Han, W. K. S. Tang, and Q. Jia, "Event-Triggered Synchronization for Nonlinear Multi-Agent Systems With Sampled Data," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, pp. 3553-3561, 2020.

[19] L. Jiang, L. Liu, J. J. Yao, L. L. Shi. "A User Interest Community Evolution Model Based on Subgraph Matching for Social Networking in Mobile Edge Computing Environments", Journal of Cloud Computing, vol. 9 no .1, 2020, pp. 69-82.

[20] A. Monney, Y. Zhan, J. Zhen, and B. B. Benuwa, "A Multi-Kernel Method of Measuring Adaptive Similarity for Spectral Clustering," Expert Systems with Applications, vol. 159, pp. 157-168, 2020.

[21] H. Lu, S. Liu, H. Wei, and J. Tu, "Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network," Expert Systems with Applications, vol. 159, pp. 113-123, 2020.

[22] J. Gou, B. Hou, Y. Yuan, W. Ou, and S. Zeng, "A new discriminative collaborative representation-based classification method via l2 regularizations," Neural Computing and Applications, vol. 32, pp. 9479–9493, 2020.

[23] J. Gou, L. Wang, B. Hou, J. Lv, Y. Yuan, and Q. Mao, "Two-Phase Probabilistic Collaborative Representation-Based Classification," Expert Systems with Applications, vol. 133, pp. 9-20, 2019.

[24] T. Guan, F. Han, and H. Han, "A Modified Multi-Objective Particle Swarm Optimization Based on Levy Flight and Double-Archive Mechanism," IEEE Access, vol. 7, pp. 183444-183467, 2019.

[25] A. Hp, A. Kz, and B. Nn, "Assessing information diffusion models for influence maximization in signed social networks - ScienceDirect," Expert Systems with Applications, vol. 119, pp. 476-490, 2019.

[26] Z. Z. Alp and u. G. üdücü, "Influence Factorization for identifying authorities in Twitter," Knowledge-Based Systems, Elsevier, vol. 163, pp. 944-954, 2019.

[27] N. Dakiche, B. S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," Information Processing & Management, Elsevier, vol. 56, no. 3, pp. 1084-1102, 2019.

[28] Y. Xiao, C. Song, and Y. Liu, "Social hotspot propagation dynamics model based on multidimensional attributes and evolutionary games," Communications in Nonlinear Science and Numerical Simulation, vol. 67, pp. 13-25, 2019.

[29] F. Figueiredo, F. Benevenuto, and J. M. Almeida, "The tube over time: Characterizing popularity growth of YouTube videos," in Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, ACM, Hong Kong, China, February 9-12, 2011: 745-754.

[30] H. L. Zhu, X. C. Yun, and Z. S. Han, "microblog popularity prediction method based on propagation acceleration," computer research and development, vol. 55, pp. 1282-1293, 2018.

[31] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," in 2010 IEEE Second International Conference on Social Computing, IEEE, Minneapolis, MN, USA, August 20-22, 2010, pp. 177-184.

[32] Y. Artzi, P. Pantel, M. Gamon. "Predicting responses to microblog posts" in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Montréal, Canada, June 3-8, 2012, pp. 602-606.

[33] M. Hasan, M. A. Orgun, and R. Schwitter, "A survey on real-time event detection from the Twitter data stream," Journal of Information Science, vol. 44, pp. 443-463, 2018.

[34] S. Dutta and A. K. Das, "A Comparative Study on Cluster Analysis of Micro-Blogging Data," in INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES IN DATA MINING AND INFORMATION SECURITY. Springer, Singapore, 2019: 873-881.

[35] H. J. Choi and C. H. Park, "Emerging topic detection in twitter stream based on high utility pattern mining - ScienceDirect," Expert Systems with Applications, vol. 115, pp. 27-36, 2019.

[36] F. Chen, A. Farahat, and T. Brants, "Multiple Similarity Measures and Source-Pair Information in Story Link Detection,". In

Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL. Boston, Massachusetts, USA, May 2-7, 2004, pp.313–320.

[37] J. Makkonen. "Investigations on event evolution in TDT, " in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Proceedings of the HLT-NAACL 2003 student research Workshop-Volume 3. ACM, Stroudsburg, PA, United States, May 3-8, 2003, pp.43-48.

[38] C. C. Yang, X. Shi, and C. P. Wei, "Discovering Event Evolution Graphs From News Corpora," IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans, vol. 39, pp. 850-863, 2009.

[39] R. Nallapati, A. Feng, F. Peng, et al. "Event threading within news topics, " in Proceedings of the thirteenth ACM International Conference on Information and Knowledge Management. ACM, Washington, D.C., USA, November 8-13, 2004, pp.446-453.

[40] P. Lee, L. Lakshmanan, E. Milios. "Cast: a context-aware story-teller for streaming social content, " in Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, Shanghai, China, November 3-7, 2014, pp.789-798.

[41] C. Lin, C. Lin, J. Li, et al. "Generating event storylines from microblogs, ". in Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, Maui, HI, USA, October 29-November 2, 2012, pp. 175-184.

[42] J. Chen, Q. Shang, H. Xiong. "Hot events detection for chinese microblogs based on the TH-LDA model", in 2018 International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018). Atlantis Press, Chengdu, China, October 31-November 1, 2018, pp.157-166.

[43] S. Mishra. "Bridging models for popularity prediction on social media" in Proceedings of the 21th ACM International Conference on Web Search and Data Mining. ACM, Melbourne VIC, Australia, Febrary 11-15, 2019, pp.810-811.

[44] D R. Liu, H. Omar, C. H. Liou, et al., " Recommending blog articles based on popular event trend analysis". Information Sciences, Elsevier, vol. 305, pp. 302-319, 2015.

[45] Dazhen, Lin, Ben, Ma, Min, Jiang, et al., "Social Network Rumor Diffusion Predication Based on Equal Responsibility Game Model," IEEE Access, IEEE, vol. 7, pp. 4478-4486, 2019.

[46] D Miao, L Liu, R Xu, J Panneerselvam, Y Wu, W Xu. ,"An efficient indexing model for the fog layer of industrial internet o f things", IEEE Transactions on Industrial Informatics 14 (10), 4487-496, 2018.

[47] P.Agarwal, R. Verma , et al. "DyPerm: Maximizing Permanence for Dynamic Community Detection." in 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2018:1-13.

[48] L. Jiang, L. Shi, L. Liu, J. Yao, B. Yuan, and Y. J. I. I. o. T. J. Zheng, "An Efficient Evolutionary User Interest Community Discovery Model in Dynamic Social Networks for Internet of People," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9226 - 9236, 2019.

[49] L.-L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy, "Event detection and user interest discovering in social media data streams," IEEE access, vol. 5, pp. 20953-20964, 2017.