Research Article

Research on Multifeature-Based Superposter Identification in Online Learning Forums

Changri Luo ¹, Xinhua Zhang ², Tingting He,³ Yong Zhang ³, Neal Xiong,⁴ and Zizhou Lu ¹

¹School of Vocational and Continuing Education, Central China Normal University, Wuhan 430079, China

²School of Computer Science Wuhan Vocational College of Software and Engineering, Wuhan 430205, China

³Academy of Computer Science, Central China Normal University, Wuhan 430079, China

⁴Northeastern State University, Department of Mathematics and Computer Science, Talequah, OK, USA

Correspondence should be addressed to Yong Zhang; 12081207@qq.com

Received 22 February 2020; Accepted 3 April 2021; Published 15 April 2021

Academic Editor: Hongju Cheng

Copyright © 2021 Changri Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of online learning and distance education, online learners' discussions in forums become increasingly effective to facilitate learning. Superposters, who play a more and more important role in forums, have attracted researchers' close attention. The key to the research is how to identify superposters among a large number of participants. Some studies focus on the network interaction of superposters and some content-related features but neglect the basic quality like language expression that a superposter should possess and the learning-related features like learning collaboration. Based on the analysis of online learning corpus, through network interaction and combination of the different features of N-gram, the paper proposed the superposter identification method based on the three primary features including language expression (L), content quality (C), and social network interaction (S) and the eight secondary features including learning collaboration. The paper applied the method in the real online learning forum corpus for identifying 28 preset superposters, achieving the results of $P@15 = 1.0$, $\text{Avg.P@15} = 1.0$, $P@28 = 0.86$, and $\text{Avg.P@28} = 0.95$. Experiments showed that this was an effective superposter identification method in online learning forums.

1. Introduction

With the improvement in online learning and remote education, discussions in online forums become increasingly effective to facilitate learning. Through online messages with teachers, learners may solve problems and ease emotional loneliness during learning. Previous research study has proved that the opinion of leaders plays an important role in online learning and has a positive effect on interactions [1, 2]. Their posts may significantly help themselves and others to learn. To differentiate between the opinion leaders in social networks, this paper terms them as superposters in learning forums. At present, research is almost nonexistent on superposter identification in online learning forums, unlike the opinion leaders in a traditional sense where much research exists [3–10]. In the context of online learning

forums, superposters refer to the users who are active in posting high-quality information, which may help learners to solve problems and prompt learning [11]. Considering the differences in discourse environment, the superposters of online learning forums differ from popular opinion leaders in social networks. Opinion leaders in social networks mainly spread information via the Internet and thus exert an influence on information receivers in terms of public opinions and tend to affect public opinions. Therefore, according to the explanation of superposters and opinion leaders, there are similarities and differences between them. Both are active in interactions; superposters aim to boost cooperative study, but opinion leaders try to influence public opinions by swaying others. How do we identify superposters among thousands of online learners? Previous similar research was made on the basis of social online

communities and applied in the fields of society and economy, but little was based on the forums of online learning platforms and applied in the field of education [5]. Although Reppel [12] asserted the application was applied in education, the research was mainly made on blogs for identifying opinion leaders in online learning communities.

Through analysis of authentic online learning forums and the characteristics of superposters, this paper obtained three points with which superposters and ordinary learners were distinguished, whereby a model framework for superposter identification in online learning forums was constructed. The framework considers both of the network interaction structure of learners and the discourse features of posts, so as to better identify superposters. Experiments showed that considering the above appropriate different features was effective in identifying superposters in online learning forums.

There are two main contributions of this work: one is proposing a new framework to identify the superposters in learning forums, and the other is proving the framework is useful for identifying the superposters in online learning forum, by experimenting on real-online learning forum corpus.

In the following, the related work will be reviewed in Section 2, the superposter identification framework will be detailed in Section 3, the experimental design and result will be analyzed in Section 4, and discussion and summary will be made in full in Section 5.

2. Related Work

Opinion leaders play a significant role in social networks. As a result, identifying opinion leaders in the context of social networks attracts the great attention of the related researchers like those in the fields of sociology and business. The role includes participation in social politics [13], promotion and popularization of new products or services in the field of business, and effect on decisions made by other consumers [6, 12]. According to the current literature, the following are main methods of identifying opinion leaders:

- (1) Identification based on network interaction structure: on the basis of the structure, in combination with users' social influence and attributes of web links, this is to reflect users' centrality and prestige in social networks through web link addresses, such as the famous PageRank, HITS algorithm, and social network analysis, which are used to identify opinion leaders [4, 6, 10, 14]. With these methods, the network interaction structure with graph models is simulated to observe the importance of user nodes, which emphasizes the structure but fails to consider the comments of opinion leaders; moreover, in the event network nodes increase for the purpose of increasing the amount of information, the graph structure will become so complicated that opinion leaders cannot be identified effectively [15].
- (2) Identification in combination with network interaction and post contents: in consideration of such limitations as sole dependence on network interaction, plenty of research is made to identify opinion

leaders in combination with post contents and network interaction. Based on social network analysis and user comments, Bodendorf and Kaiser explored the opinion leaders in online communities and the propagation trend of the public opinions they make [16]. In combination with the features of network structure and user behavior and the emotional features of posts, through analysis of multidimensional features, Cao et al. studied the social network-based opinion leaders [17]. Li and Du constructed an opinion leader identification framework with blog contents, author attributes, reader attributes, and the network relationship between blog authors and readers to identify the opinion leaders committed to word-of-mouth marketing in online social blogs [5]. Although good results were achieved, the above research depended too much on influence or centrality, making it impossible to reflect the quality of the contents published by opinion leaders and thus accurately identify opinion leaders. Meanwhile, they were made based on social networks instead of identification in the field of education. In accordance with the features of expertise, novelty, influence, activity, longevity, and centrality, Li and Ma et al. built an indicator framework to identify opinion leaders [18]. Huang et al. identified superposters according to the quantity and quality of learners' posts in course forums [19]. This is rare in terms of opinion leader (superposters) identification in online learning forums but fails to reflect the quality and role in cooperative study of superposters' posts.

In the opinion of the author, superposters in learning forums are different from opinion leaders in social network, and they must have a certain cultural quality and cooperative study skills, which are not reflected in the above research studies. Therefore, in consideration of the limitations of the abovementioned studies on opinion leader identification, this paper proposes a superposter identification framework based on language expression, content quality, and interaction structure, so as to identify superposters among the participating learners and learning supporters.

3. Superposter Identification Framework

The authors consider that the superposters in online learning forums should be as follows: (1) be active in posting/replying; (2) be excellent in language expression; (3) post high-quality posts and have a good ability to learn, or be knowledgeable, or accurately reflect learning needs, or provide other assistances to online learners. These not only reflect the importance of poster nodes in interaction through forums but also indicate the authority of their posts. Based on these features, the paper proposes the framework (see Table 1) for superposter identification in online learning forums (Chinese as the working language), as shown in Table 1. According to the definition given by this paper, for a superposter, we expect to reflect the language expression level of learners, quality of post contents, and activity of interaction, respectively, through language expression, content quality, and social network interaction.

TABLE 1: Indexes and description of model features.

Items	Features (symbol)	Description
Language expression (Le_Index)	Word normalization (W_I)	To survey the use of Class I and Class II Chinese characters in posts
	Term nonnormalization (T_I)	To survey the use of uncivil words and Internet slangs in posts
	Language elegance (Le_I)	To survey the use of words, phrases, and idioms
	Learning collaboration (Lc_I)	To survey posters' ability to solve others' problems
Content quality (Cq_Index)	Correlation with the thread (Ct_I)	To survey the similarity between the reply content and the thread title
	Expertise of content (Ec_I)	To survey the knowledge points involved in posts
Social network interaction (Sns_Index)	Out-degree centrality (DC_o)	To survey the intermediate status of posters in network structure
	In-degree centrality (DC_i)	To survey the authority of posters in network structure

3.1. Social Network Interaction. In social network analyses, degree centrality is an important index that measures the social interaction of individuals as well as a common index that evaluates the social status and prestige of individuals, including out-degree centrality and in-degree centrality; out-degree centrality is used to reflect the replies of a poster (learner or learning supporter) a_i to others' posts, as expressed with the following formula:

$$DC_o(a_i) = \frac{\text{NumReplyOut}(a_i)}{N-1} \times \frac{\sum_{j=1, j \neq i}^N a_{ij}}{\sum_{i=1}^N \sum_{j=1}^N a_{ij} - \sum_{i=1}^N a_{ii}}, \quad (1)$$

where a_i is the i^{th} learner of learners set A ; N is the total number of learners, similarly hereinafter; $\text{NumReplyOut}(a_i)$ is the reply of a_i to others, i.e., the number of linkout of node a_i in interactive networks, which reflects the importance of node position; $\sum_{j=1, j \neq i}^N a_{ij} / (\sum_{i=1}^N \sum_{j=1}^N a_{ij} - \sum_{i=1}^N a_{ii})$ is the ratio of the number of a_i 's replies to others' posts to the total number of replies (excluding all self-replies), which reflects the degree of interaction in which a_i participates; and traditional algorithms only consider $\text{NumReplyOut}(a_i) / (N-1)$ but ignores the degree of interaction reflected by $\sum_{j=1, j \neq i}^N a_{ij} / (\sum_{i=1}^N \sum_{j=1}^N a_{ij} - \sum_{i=1}^N a_{ii})$.

Degree centrality, also known as Prestige [20], may reflect the replies of other posters to the posts of a_i , as expressed with the following formula:

$$DC_i(a_i) = \frac{\text{NumReplyIn}(a_i)}{N-1} \times \frac{\sum_{j=1, j \neq i}^N a_{ji}}{\sum_{j=1}^N \sum_{i=1}^N a_{ji} - \sum_{i=1}^N a_{ii}}, \quad (2)$$

where a_i is the i^{th} learner of set A ; N is the total number of learners; $\text{NumReplyIn}(a_i)$ is the number of link in of node a_i in interactive networks, which reflects the node prestige; and $\sum_{j=1, j \neq i}^N a_{ji} / (\sum_{j=1}^N \sum_{i=1}^N a_{ji} - \sum_{i=1}^N a_{ii})$ is the ratio of the total

number of others' replies to a post of a_i to the total number of others' replies (excluding all self-replies), which reflects the centrality of a_i in interactive networks but is rarely considered in traditional algorithms.

Therefore, the index of social network interaction of a_i is calculated as follows:

$$\text{Sns_Index}(a_i) = \sigma \times DC_o(a_i) + (1 - \sigma) \times DC_i(a_i), \quad (3)$$

where σ is a weighting parameter.

3.2. Language Expression. Plenty of research on identification of opinion leaders failed to consider the language expression skill of an opinion leader. However, whether in terms of interaction in social networks or online learning forums, an opinion leader or a superposter must ensure fluent language expression and owns a certain cultural quality. If a post involves violent words or unclear expressions all along, no matter how innovative or important it is, other users (learners of learning forums) may refuse to discuss further. For this reason, this paper makes a survey on language expression with three indexes including "word normalization," "term nonnormalization," and "language elegance." These are relatively easily achieved and may reflect the language expression skill of posters.

3.2.1. Word Normalization. Word normalization is to survey the frequency of Class I and Class II commonly used Chinese characters in posts and thus verify the normalization of the words used by learners. When uncommon words are used in posts to appear intellectual, learners may find it difficult to achieve optimal learning, thus limiting the spread of information. To facilitate survey, the index of the normalization of the words used by a_i is defined as follows:

$$W_I(a_i) = 0.9 \times \frac{\text{CH_I_Freq}(a_i)}{\text{Total_CH_Freq}(a_i)} \times \frac{\text{CH_I_Type}(a_i)}{2500} + 0.1 \times \frac{\text{CH_II_Freq}(a_i)}{\text{Total_CH_Freq}(a_i)} \times \frac{\text{CH_II_Type}(a_i)}{1000}, \quad (4)$$

where $\text{CH_I_Freq}(a_i)$ and $\text{CH_II_Freq}(a_i)$, respectively, are the frequency of Class I and Class II commonly used Chinese characters in all posts of a_i ; $\text{Total_CH_Freq}(a_i)$ is the total

frequency of Chinese characters in all posts of a_i ; and $\text{CH_I_Type}(a_i)$ and $\text{CH_II_Type}(a_i)$, respectively, are the number of the types of Class I and Class II Chinese

characters in all posts of a_i . Constants 2500 and 1000, respectively, are the number of the types of Class I and Class II Chinese characters; 0.9 and 0.1 are weighting parameters and empirical values.

3.2.2. Term Nonnormalization. Term nonnormalization is to survey the use of uncivil words by learners (Internet users). Such usage involves impolite, violent, and vulgar words and some Internet slangs in the process of exchange in forums. Thus, to further analyze the normalization of the words use by surveying the use of uncivil words and Internet slangs, the paper defines the index of term nonnormalization of a_i as follows:

$$T_I(a_i) = \frac{C \times \text{UnC_W_Freq}(a_i) + \text{Net_W_Freq}(a_i)}{\text{TotalWFreq}(a_i)}, \quad (5)$$

where $C = 1.2$ as a constant and reflects that uncivil words are more improper than Internet slangs and $\text{UnC_W_Freq}(a_i)$, $\text{Net_W_Freq}(a_i)$, and $\text{TotalWFreq}(a_i)$, respectively, are the frequency of uncivil words and frequency of Internet slangs in all posts of a_i and the total frequency of words.

3.2.3. Language Elegance. Language elegance is to survey the use of fixed phrases (including fixed terms, phrases, and idioms) in posts. Although language derives from life, we cannot deny the fact that “individualized teaching” (yin cai shi jiao-因材施教, Chinese idiom) is more concise, refined, and elegant than “adopting different teaching methods for different students” in terms of expression. If similar expressions are frequently used in a post, we may see the vocabulary and language mastery of the poster. Accordingly, the paper observes the language expression ability based on this. The language elegance of a_i is calculated as follows:

$$\text{Le}_I(a_i) = \sigma_1 \times \frac{\text{CW 1 Freq}(a_i)}{\text{Total W Freq}(a_i)} \times \frac{\text{CW 1 Type}(a_i)}{\text{Type 1 Num}} + \sigma_2 \times \frac{\text{CW 2 Freq}(a_i)}{\text{Total W Freq}(a_i)} \times \frac{\text{CW 2 Type}(a_i)}{\text{Type 2 Num}} + (1 - \sigma_1 - \sigma_2) \times \frac{\text{Idioms Freq}(a_i)}{\text{Total W Freq}(a_i)} \times \frac{\text{Idioms Type}(a_i)}{\text{Type I Num}}, \quad (6)$$

where $\text{CW1 Freq}(a_i)$, $\text{CW2 Freq}(a_i)$, $\text{Idioms Freq}(a_i)$, and $\text{Total W Freq}(a_i)$ are the frequency of commonly used class I words, frequency of commonly used class II words, frequency of idioms in all posts of a_i , and the total frequency of words, respectively; $\text{CW1Type}(a_i)$, $\text{CW2Type}(a_i)$, and $\text{Idioms Type}(a_i)$, respectively, is the number of the types of Class I and Class II words and idioms;

Type 1 Num, Type 2 Num, and Type I Num, respectively, is the total number of the types of Class I and Class II words and idioms; and $\sigma_1 = 0.25$ and $\sigma_2 = 0.35$ as constants, which are the coefficients from locally optimal solutions obtained through repeated experiments and empirical values.

Therefore, the index of language expression of a_i is calculated as follows:

$$\text{Le_Index}(a_i) = \vartheta_1 \times W_I(a_i) - \vartheta_2 \times T_I(a_i) + (1 - \vartheta_1 - \vartheta_2) \times \text{Le}_I(a_i), \quad (7)$$

where ϑ_j as weighting parameters, $j = 1, 2$.

3.3. Content Quality. The content quality of posts directly affects the result of interaction in online learning forums. Therefore, in the process of identifying superposters, surveying the content quality is very important. In surveying the quality of the contents posted by superposters, the paper mainly focuses on three questions:

- (i) Q1: do the post contents help others to solve learning problems?
- (ii) Q2: are replies relevant to a topic?
- (iii) Q3: how about the conformity of post contents with knowledge points?

That is, if a_i is a superposter, his/her posts will be considered high-quality, helpful to others in the process of interaction and highly relevant to a topic (rather than spam or meaningless posts) and to have highly professional knowledge points. Therefore, the paper evaluates content

quality based on the learning collaboration, correlation with the thread, and expertise of content.

3.3.1. Learning Collaboration. Learning collaboration, mainly used to observe the role of posts and interaction activities in supporting participants to learn, is to survey whether post contents may help others to solve problems herein. The learning collaboration of a_i is defined as follows:

$$\text{Lc}_I(a_i) = \frac{\text{HelpPostNum}(a_i)}{\text{TotalPostNum}(a_i)} \times \frac{\text{BeneficiaryNum}(a_i)}{N}, \quad (8)$$

where $\text{HelpPostNum}(a_i)$, $\text{TotalPostNum}(a_i)$, and $\text{BeneficiaryNum}(a_i)$, respectively, is the number of helpful posts of a_i , total number of posts, and the number of the beneficiaries from the posts of a_i .

The present difficulty is how to confirm whether a post of a_i may help others to solve problems. There is also similar research, including that on manual confirmation, which is

time and labour consuming and undesirable for massive corpus, and that on automatic confirmation, which identifies answer-question in forums in combination with rules and forum structure and achieves good results in extraction experiments [21]. In addition, this is also confirmed by vote in forums [22]. With the second method, the paper confirms the data about helpful posts and beneficiaries in line with rules and in a statistical manner.

3.3.2. Correlation with the Thread. Correlation with the thread title is the correlation of replies to the topic discussed in the main post. In the process of discussion in online learning forums, learners often post some irrelevant comments about question B in a post where question A is discussed. A superposter must not or rarely do so and should

comment in accordance with threads. Therefore, the correlation of a reply of a_i to topic is calculated as follows:

$$Ct_I(a_i) = \frac{\text{CorrNum}(a_i)}{\text{TotalReplyNum}(a_i)}, \quad (9)$$

where $\text{CorrNum}(a_i)$ and $\text{TotalReplyNum}(a_i)$, respectively, are the number of the replies of a_i considered relevant to the target topic (main post) and the total number of the replies of a_i . The correlation between a reply of a_i and the target thread is calculated with the cosine value.

3.3.3. Expertise of Content. Expertise of content is the course knowledge points involved in a post, by which the index of expertise of content in the posts of a_i in the C_j forum can be calculated as follows:

$$Ec_I(a_i) = \frac{\text{KnowledgePointNum}(a_i, C_j)}{\text{TotalKnowledgePointNum}(C_j)} \times \frac{\text{KnowledgePostNum}(a_i, C_j)}{\text{TotalPostNum}(a_i, C_j)}, \quad (10)$$

where $\text{KnowledgePointNum}(a_i, C_j)$, $\text{KnowledgePostNum}(a_i, C_j)$, and $\text{TotalPostNum}(a_i, C_j)$, respectively, are the frequency of the knowledge points of C_j included in the posts posted by a_i , number of the posts which contain at least 1 knowledge point, and the total number of the posts posted in the C_j forum. $\text{TotalKnowledgePointNum}(C_j)$ is the total

frequency of knowledge points of C_j in the forum. Each knowledge point which appears for 1 or 0 times is not counted repeatedly. $Ec_I(a_i)$ is the index of educational content of the posts sent by a_i in the C_j ($j = 1, \dots, K$) forum.

Accordingly, the content quality of a_i can be calculated as follows:

$$Cq_Index(a_i) = \mu_1 \times Lc_I(a_i) + \mu_2 \times Ct_I(a_i) + (1 - \mu_1 - \mu_2) \times Ec_I(a_i), \quad (11)$$

where μ_i , ($i = 1, 2$) are weighting parameters.

3.4. Superposter Index. With the MIN-MAX method, the paper normalizes the results of Le_Index , Cq_Index , and Sns_Index . For example, Le_Index can be normalized with the following formula to an extent that realizes the result within the range from 0 to 100:

$$NLe_Index(a_i) = \frac{\text{Le_Index}(a_i) - \text{MIN}(\text{Le_Index}(a_j)_{j=1}^N)}{\text{MAX}(\text{Le_Index}(a_j)_{j=1}^N) - \text{MIN}(\text{Le_Index}(a_j)_{j=1}^N)} \times 100. \quad (12)$$

Similarly, we may obtain NCq_Index and $NSns_Index$.

In conclusion, superposter index ($Super_I$) is calculated as follows:

$$Super_I = (1 - \alpha - \beta) \times NLe_Index + \alpha \times NCq_Index + \beta \times NSns_Index, \quad (13)$$

where α and β are weighting parameters which may be set according to the actual situation.

So Algorithm 1 can be described as ALG_Super_1 (see Algorithm 1, ALG_Super_1).

4. Experimental Result and Analysis

4.1. Data Set. The data in the paper are downloaded from the Q&A forum [23] for online learning course Computer Application Foundation. The dataset includes 7494 subject, 22369 posts, and 6747 participants (including 6712 learners

and 35 teachers). Among the 35 teachers, 28 were found to meet the defined conditions of a superposter, through sampling and analysis of the data about their posts (see Table 2). Therefore, 28 teachers were considered as superposters and identified with the method proposed in the paper; there were 7 teachers unqualified to be superposters for 4 teachers who posted 2 posts each and 3 teachers who posted 1 post each. In addition, to count the number of knowledge points in posts, the paper constructs an online unified examination knowledge point set based on the Fundamentals of Computer Application [24].

Input: the post dataset of online learners, D ; training parameter set, P .
Output: Superposter index, Super_I .

- (1) $A \leftarrow$ Counting online learners.
- (2) For each online learner a_i
 - (1) Computes $\text{Sns_Index}(a_i)$ by using formula (4);
 - (2) Computes $\text{Le_Index}(a_i)$ by using formula (7);
 - (3) Computes $\text{Cq_Index}(a_i)$ by formula (11);
 - (4) With the MIN-MAX method, normalizes the results of $\text{Le_Index}(a_i)$, $\text{Cq_Index}(a_i)$ and $\text{Sns_Index}(a_i)$:
 $\text{NSns_Index}(a_i) \leftarrow \text{Sns_Index}(a_i)$;
 $\text{NCq_Index}(a_i) \leftarrow \text{Cq_Index}(a_i)$;
 $\text{NLe_Index}(a_i) \leftarrow \text{Le_Index}(a_i)$;
 - (5) Computes superposter index $\text{Super_I}(a_i)$ by using formula (13).
- (3) Rank Super_I ;
- (4) Return Super_I .

ALGORITHM 1: ALG_Super_1: framework of computing superposter index.

4.2. Evaluation. There have been no mature and recognized methods of assessing superposter identification. In this section, evaluation is made with the following indexes [11].

The accuracy of TOP M ($P@M$) is as follows:

$$P@M = \frac{\text{the number of correct super posters in TOP } M}{M}. \quad (14)$$

The average accuracy of TOP M ($\text{Avg } P@M$) is as follows:

$$\text{Avg } P@M = \frac{\sum_{i=1}^M P@i}{M}. \quad (15)$$

4.3. Result Analysis. With the three feature indexes of the model including language expression (L), content quality (C), and network interaction structure (S), the paper makes a test on the effect of identifying the 28 superposters. Through repeated weighting tests on data, the weighting parameters are set in Table 3, and the results and statistical analysis are described in Tables 4 and 5, respectively. Comparison of the results of our algorithm with the PageRank algorithm (PR) is shown in Tables 5 and 6 and Figures 1–3.

According to Table 5, ① the model indexes achieve good results of superposters identification, and the model is very effective in application in the dataset. ② In terms of the identification result achieved by each feature, content quality, which realizes the average accuracy of over 0.9, is considered best. Language expression is just as the N-gram model, in which only 14 superposters are correctly identified among TOP 28, and is considered good. Social network structure, a common index in social network analysis by which 22 are correctly identified among TOP 28, is considered better. ③ Although a single feature is unable to perform well, their combinations may realize striking effects: with the combinations like LC, CS, and LCS, 24 superposters are correctly identified among TOP 28. With the combinations like LC and LCS, all of the superposters can be identified among TOP 15,

which undoubtedly proves that the feature designs are rational and effective. ④ The single feature L performs averagely, but its combination with other features performs well, especially LC. We are confused about whether this means that the two models are mutually complementary as part of contents in terms of structure. ⑤ Among TOP 28, 14 superposters are identified by L and LS models. This shows that language expression greatly depends on the length of text and is not sensitive to identification of the superposters of short text although consideration has been given to avoidance of this case in design. Since the number of the superposters identified by LS is less than that by S, we are confused about whether this means that L and S have something in common. However, there is a difference between L as a content-based result and S as a graph-based (network interaction) result in structure. We are confused that whether it is or not because the two models with different structures are mutually exclusive, which causes the result to deteriorate. Actually, this is also the case with the CS model, which achieves a result inferior to that C does, which we cannot explain in this study. Particularly, in the event of TOP 51, all of the 28 superposters can be identified. From Table 5, the trend chart for identification results and for average identification results achieved by each feature in different cases can be obtained (see Figures 1 and 2). Compared with the PageRank algorithm, the experiment result (see Tables 5 and 6 and Figures 1–3) of our algorithm (LCS) is better.

4.4. Discussion. In social networks, it is widely believed that out degree is as important as in degree; in the process of testing the weighting parameter of S, we found that when $\sigma = 1$, a good result was achieved locally and other values slightly improved; however, this was considered very unbalanced; that is, it only considered the replies to others' posts but neglected others' replies to own posts, leaving it not universal; through observation of data set, we found that there is a difference between the number of threads and the number of replies, especially in relation to the 28 teachers; in consideration of the generality of the model, through

TABLE 2: Comparison between two kinds of user of different identities in basic indexes.

Objects	Maximum number of posts sent	Maximum number of replies	Less than 10 posts	Average number of posts sent	Total number of Chinese characters in posts	Total number of words in posts
Teacher (35)	8617	8565	10	307.86	275340	147897
Learner (6712)	47	25	6658	1.73	548607	284833

TABLE 3: Parameters settings.

Parameters	Value
ϑ_1	0.2
ϑ_2, σ	0.5
μ_1	0.3
μ_2	0.4

TABLE 4: TOP 10 examples in L , C , and S .

User ID	L	C	S	Super_1
1527315	96.91504829	99.99880313	99.99430712	99.3807
1196662	98.79587171	53.61549801	0.868037451	46.82733
1191173	99.99825778	52.56897262	0.407785774	46.40647
1031382	97.07805875	49.47187606	0.023758924	44.15868
1191171	95.15300193	49.2480221	0.070274765	43.67569
1348	94.43927005	48.98495494	0.027054183	43.38845
511546	94.76087703	48.63108673	0.007664835	43.27002
1149	93.84371959	48.82328467	0.019282605	43.18617
1093505	92.56949365	48.49936007	0.020990493	42.76988
1237	93.38233965	48.13972648	0.002324243	42.74703

TABLE 5: Identification of superposters with each feature and their combinations.

Features	P@5	Avg P@5	P@10	Avg P@10	P@15	Avg P@15	P@20	Avg P@20	P@28	Avg P@28
L	0.8	0.96	0.7	0.80	0.6	0.74	0.55	0.69	0.5	0.65
C	1.0	1.0	0.9	0.94	0.93	0.93	0.9	0.93	0.86	0.92
S	0.8	0.91	0.9	0.89	0.87	0.88	0.9	0.88	0.79	0.86
LC	1.0	1.0	1.0	1.0	1.0	1.0	0.95	0.9975	0.86	0.98
LS	1.0	1.0	0.7	0.85	0.6	0.78	0.55	0.72	0.5	0.70
CS	0.8	0.96	0.9	0.92	0.93	0.92	0.9	0.92	0.86	0.91
LCS	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.98	0.86	0.95
PR	1	1	0.8	0.97	0.86	0.93	0.8	0.89	0.64	0.84

Note. $LC = 0.4L + 0.6C$, $LS = 0.3L + 0.7S$, $CS = 0.6C + 0.4S$, $LCS = 0.2L + 0.45C + 0.35S$.

TABLE 6: Recall and accuracy of identifying 28 superposters by LCS and PR.

	TOP 5	TOP 10	TOP 15	TOP 20	TOP 28	TOP 51	TOP 571
Recall-LCS	0.179	0.357	0.5367	0.64	0.86	1	—
Accuracy-LCS	1	1	1	0.9	0.86	0.549	—
Recall-PR	0.179	0.286	0.464	0.571	0.643	0.786	1
Accuracy-PR	1	0.8	0.86	0.8	0.643	0.43	0.049

comprehensive consideration, the paper sets $\sigma = 0.5$. Other parameters are set according to the optimum effects achieved by a single feature in experiments.

According to the intermediate results based on content quality, we found that some data were lost, such as the number of beneficiaries and helpful posts, especially as to learning supporters; for example, the posts sent or replies by teachers were related to the course and helpful to learners; therefore, all of the learners who participate in the

interaction were beneficiaries, and the posts of the teachers were considered helpful; however, the paper was unable to accurately obtain such information, leading to a loss of related data and affecting the identification of the teachers as superposters (the relationship between recall and accuracy can be seen in Table 6 and Figure 3). Nevertheless, through the model (LCS), all of the 28 superposters can be identified among TOP 51. It can be done by PageRank algorithm at top 571 cases.

FIGURE 1: Trend chart for identification results achieved by each feature in different cases.

FIGURE 2: Trend chart for average identification results achieved by each feature in different cases.

FIGURE 3: Trend chart of recall of and accuracy of identifying 28 superposters by LCS and PR.

5. Conclusion

Through analysis of the data on the posts sent by learners in online learning forums, the paper proposed a superposter identification model based on characters, words, and network interaction structure. First, through analysis of the network interaction of users based on graph structure, the paper calculated the out-degree centrality and in-degree centrality of each user node in networks, which involved both the interaction breadth and depth of each node, so as to determine its activity and importance in interactive networks. Then, learners' language expression was included in the identification framework, including word normalization, term normalization, and language elegance, by which the normalization of the words and terms used by learners and their basic ability to master language are judged. The third-dimensional feature is most important in online learning forums, i.e., content quality, which includes learning collaboration, correlation with the thread, and expertise of content. An online learning forum is designed to facilitate cooperation between learners and interaction in relation to learning contents. Learning collaboration mainly considers whether a post is helpful to others in study; correlation with the thread is to verify the correlation between a post and the topic discussed therein; expertise of content is to survey whether course knowledge points are included in a post. Accordingly, the three indexes work, respectively, in online learning forums on a targeted basis.

Although there are some deficiencies in the design of the superposter identification model, such as the need of repeated experiments on manual setting of weighting parameters in the process of calculation, a good result was achieved with the proposed method for identifying the preset 28 superposters. Considering that the method is easily realized and involves few calculations, and it is worthy to be applied in practical online learning systems.

Data Availability

Research data related to learners' personal privacy cannot be shared.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by China Scholarship Council and a general project of the National Natural Science Foundation of China (Event-Based Semantic Research on Educational Texts, Grant no. 61977032).

References

- [1] B. D. Wever, H. V. Keer, T. Schellens et al., "Roles as a structuring tool in online discussion groups: the differential impact of different roles on social knowledge construction," *Computers in Human Behavior*, vol. 26, no. 4, pp. 516–523, 2010.
- [2] S. Zha and C. Lee Ottendorfer, "Effects of peer-led online asynchronous discussion on undergraduate students' cognitive achievement," *American Journal of Distance Education*, vol. 26, no. 4, pp. 238–253, 2002.
- [3] N. Matsumura, Y. Ohsawa, and M. Ishizuka, "Identifying opinion leaders in the blogosphere," 2002.
- [4] X. Song, Y. Chi, K. Hino et al., "Mining and characterizing opinion leaders from threaded online discussions," 2007.
- [5] F. Li and T. C. Du, "Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs," *Decision Support Systems*, vol. 51, no. 1, pp. 190–197, 2011.
- [6] Y. Cho, J. Hwang, and D. Lee, "Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach," *Technological Forecasting and Social Change*, vol. 79, no. 1, pp. 97–106, 2012.
- [7] W. Zhang, H. He, and B. Cao, "Identifying and evaluating the internet opinion leader community based on k-clique clustering," *Neural Computing and Applications*, vol. 25, no. 3–4, pp. 595–602, 2014.
- [8] N. Ma and Y. Liu, "SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1357–1368, 2014.
- [9] P. Jia, A. Mirtabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," *Siam Review*, vol. 57, no. 3, pp. 367–397, 2015.
- [10] L. Yang, Y. Qiao, Z. Liu, J. Ma, and X. Li, "Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm," *Soft Computing*, vol. 22, no. 2, pp. 453–464, 2018.
- [11] X.-H. Fan, J. Zhao, F. A. N. G. Bin-Xing et al., "Influence diffusion probability model and utilizing it to identify network opinion leader," *Chinese Journal of Computers*, vol. 36, no. 2, pp. 360–367, 2013.
- [12] A. E. Reppel, I. Szmigin, and T. Gruber, "The iPod phenomenon: identifying a market leader's secrets through qualitative marketing research," *Journal of Product & Brand Management*, vol. 15, no. 4, pp. 239–249, 2006.
- [13] I. Himmelboim, E. Gleave, and M. Smith, "Discussion catalysts in online political discussions: content importers and conversation starters," *Journal of Computer-Mediated Communication*, vol. 14, no. 4, pp. 771–789, 2009.
- [14] Z. Zhai, H. Xu, and P. Jia, "Identifying opinion leaders in BBS," 2008.
- [15] D. K. Kim, A. C. James, and G. J. Shepherd, *Identifying Opinion Leaders by Using Social Network Analysis: A Synthesis of Opinion Leadership Data Collection Methods and Instruments*, The Faculty of the Scripps College of Communication of Ohio University, Columbus, OH, USA, 2007.
- [16] F. Bodendorf and C. Kaiser, *Detecting Opinion Leaders and Trends in Online Communities*, IEEE International Conference on Digital Society, New York, NY, USA, 2010.
- [17] J.-X. Cao, C. H. E. N. Gao-jun, W. U. Jiang-lin et al., "D multi-feature based opinion leader mining in social networks," *Acta Electronica Sinica*, vol. 44, no. 4, pp. 898–905, 2016.
- [18] Y. Li, S. Ma, Y. Zhang, R. Huang, and F. Kinshuk, "An improved mix framework for opinion leader identification in online learning communities," *Knowledge-Based Systems*, vol. 43, no. 2, pp. 43–51, 2013.
- [19] J. Huang, A. Dasgupta, A. Ghosh et al., "Superposter behavior in MOOC forums," 2014.
- [20] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, "Predicting students' final performance from participation in

- on-line discussion forums,” *Computers & Education*, vol. 68, pp. 458–472, C, 2013.
- [21] B. Wang, B. Liu, C. Sun et al., “Extracting Chinese question-answer pairs from online forums,” *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 23, 2009.
- [22] P. Jurczyk and E. Agichtein, “Discovering authorities in question answer communities by using link analysis,” 2007.
- [23] <http://zjy.ccnu.edu.cn/jyfw/wljy.htm>.
- [24] National College Network Education Examination Committee Office, *Fundamentals of Computer Application*, Tsinghua University Press, London, UK, 2013.