Research Article

English News Text Recommendation Method Based on Hypergraph Random Walk Label Expansion

Zhehao Wang

Department of Culture Education, Henan Institute of Economics and Trade, Zhengzhou, Henan 450000, China

Correspondence should be addressed to Zhehao Wang; 81467@hnuhe.edu.cn

Received 9 January 2022; Revised 11 February 2022; Accepted 17 February 2022; Published 18 March 2022

Academic Editor: Gopal Chaudhary

Copyright © 2022 Zhehao Wang. 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 rapid development of multimedia and Internet technology, English news text summary technology has received widespread attention as a way to quickly obtain English news text content. The existing method of English news text summarization based on graph model usually takes English news text as the vertices of the graph, and the relationship between the two vertices is represented by the edge. Although it has achieved good results, it cannot quickly sort out the English complex relationships between news texts. In order to solve this problem, this paper uses the hypergraph model to model the relationship between English news text, and conducts in-depth research on the application of the hypergraph model in the field of English news text summarization. The high popularity of the Internet has brought about earth-shaking changes to the news industry, which makes the news on the Internet a great way for netizens to get news. However, the public cannot pick out satisfactory events from a large amount of news. In order to solve this problem, news event discovery technology that can help users quickly discover and understand hot news is produced. In addition, user personalized recommendation technology will rely on customer operation habits to provide customers with hot events of interest. The personalized news recommendation method adopted in this article has the advantages of integrating discovery and personalized news recommendation, and provides users with a better experience. Compared with the traditional hierarchical clustering algorithm, the algorithm proposed in this paper significantly improves the accuracy.

1. Introduction

This paper adopts a new algorithm of mixed recommendation news. This new algorithm uses two methods to conduct experiments to understand the user’s interest model [1]. First, understand the multiple characteristics of the event and combine the user’s browsing process to form a new model [2]. Secondly, build a model of the news headlines that the user has browsed, and then find groups with similar similarities for users, and then adopt a model that combines SVM and LDA models [3]. Finally, provide news users with news obtained after a series of modeling and calculations. What will appear in the user’s hands will be news that is very close to the user’s psychology and can quickly satisfy the user’s reading experience on time [4]. Based on the needs of the real society, a personalized news event recommendation system is developed in this article [5]. Empirical research shows that the system can achieve the purpose of discovering and recommending news events for users according to preset functions at a stable and efficient operating speed [6]. The first is the proposal of an English news text summarization algorithm based on hypergraph sorting. The HGRVS algorithm includes four parts: extracting features of English news text, constructing hypergraphs, classifying text, and generating summaries. The core of the HGRVS algorithm is to first use the hypergraph sorting algorithm to classify English news text, and then select the items that need to be used from the already classified categories for backup, and compare the selected items above with other news relevance. And remove redundant parts to maintain complete coverage of news events [7]. Secondly, an English news text summarization algorithm based on hypergraph random walk is proposed. This algorithm calculates the stable distribution probability of each English news text through
random walks, so as to achieve the purpose of ranking the
importance of English news texts [8]. The more important
the English news text, the greater the probability of its stable
distribution. Then referring to the above conclusions, you
need to select the English news text with the highest ranking
to perform the de-redundancy operation, and then get the
static English news text summary. Finally, after a large
number of experiments and a large number of comparisons,
the results further verify the effectiveness and advancement
of the algorithm [9]. The first method in this article builds a
user’s interest model based on the user’s behavior and habits
and integrates the multiple characteristics of the event. The
second method is to build a new model based on the news
headlines the user has browsed to quickly obtain the user’s
reading preferences [10].

2. Related Works

Literature [11] shows that the Internet and informatization
have gradually entered people’s lives. Similarly, among the
many ways people obtain information, the network has be-
come particularly important. In the world of network in-
formation explosion, the problem of information overload is
becoming more and more serious, making it impossible for
network users to quickly and accurately find information that
is of real use to them in the vast information sea. Literature
[12] shows that even though scholars have proposed three
methods to solve the problem of information overload, search
engines, portal websites and classified websites, they cannot
solve the user’s problem. Literature [13] shows that current
users are facing the problem of finding answers to questions
and obtaining recommended content of interest. The emerg-
ence of keywords provided by search engines provides users
with directions for finding answers to questions, but a few
simple keywords. But it is difficult to bear the user’s complete
idea, so the blind search by the user leads to inefficient
browsing. Literature [14] shows that as of February 2020,
China’s Internet penetration rate has reached 79.8%, and the
number of Chinese Internet users has increased significantly.
The number of Internet users in 2020 alone will reach 39.75
million. Compared with previous years, the growth rate of
Internet users has increased significantly. This shows that the
number of users who obtain information by browsing the web
has increased rapidly, but these users also face challenges and
problems. Nowadays, the mixed news information makes it
difficult for Internet users to find articles of interest. Some-
times due to the uneven level of media resources behind the
operation of the news website, the focus of website reporting
is also different, and some focus on sports events. Some
reports are more focused on financial and social aspects.
Literature [15] shows that users, as information capturers, will
not only pay attention to news in a certain industry. The
even distribution of website resources causes users to waste
time on spanning websites and fail to find all they want to see
at one time. From another perspective, most users will choose
to read this article after looking for similar news recom-
manded at the end of the article, and users will spend a lot of
time searching for such articles. For such problems, incident
discovery technology is a powerful solution. This technology
will help users sort and classify news, and then find hot events
during the period. Literature [16] shows that users can use the
saved manual search time to read more articles of interest.
Because news has strong effectiveness, the method of static
product recommendation on social and business websites
cannot meet the two requirements of news popularity and
novelty. And for some users, sometimes the public hot events
recommended by the event discovery technology cannot
catch the user’s attention, so you need to provide users with all
the content they want to learn about according to their
personal interests. And related content personalized rec-
ommendation technology. Literature [17] shows that the
news events browsed by users in different time periods are
related, so it is necessary to improve the personalized design,
so that it can promptly eliminate news events that users have
read in a short time. Provide users with new hot events at any
time. The same problem that needs to be considered in the
design is that when there are a large number of visits and news
events, the personalized recommendation technology can
take a major responsibility, so the development of the per-
formance of the personalized event recommendation system
is also facing greater challenge.

3. Label Expansion Method Based on
Hypergraph Random Walk

3.1. Basic Concepts of Hypergraph. In the field of multimedia,
a graph model can be used to represent similar things, and
two similar images can be connected together to be used in
the field of image retrieval. The graph model can describe
related things well. But there is a many-to-many relationship
in real life. For example, the same author can have multiple
papers, and the same paper can have multiple authors. At
this time, if the graph model is used to establish connections,
the paper is the vertex of the graph. When an author holds
two articles, the two articles can be connected by one side.
Hypergraph is usually represented by the incidence matrix
HVE, which is defined as follows:

\[ h(v_i, e_j) = \begin{cases} 1, & \text{if } v_i \in e_j, \\ 0, & \text{if } v_i \notin e_j. \end{cases} \tag{1} \]

The relationship between 0 and 1 is used to determine
whether a vertex belongs to a hyperedge, and the vertex plays
an extremely important role in it. Although the 0, 1 model is
simple and easy to use, it causes the problem of missing
vertices. Therefore, the probability hypergraph model is used
to solve this problem, and its incidence matrix \( H \) is defined as:

\[ h(v_i, e_j) = \begin{cases} A(i, j), & \text{if } v_i \in e_j, \\ 0, & \text{if } v_i \notin e_j, \end{cases} \tag{2} \]

Where \( A(i, j) \) is defined as

\[ A(i, j) = e^{-\text{dis}(v_i, v_j)}. \tag{3} \]

It can be seen that the emergence of hypergraph
probability model improves the probability of local infor-
mation and hyperedges. In this way, the relationship
between the vertices can be better expressed. In addition, the weight of the superedge is defined as
\[ w(e_i) = \sum_{v_j \in e_i} A(i, j). \]  
(4)

The overedge weight matrix \( W_e \) is defined as:
\[ W_e = \begin{bmatrix} w(e_1) \\ w(e_2) \\ \vdots \\ w(e_n) \end{bmatrix}. \]  
(5)

The degree of superedge is defined as:
\[ \sigma(e) = \sum_{v \in e} h(v, e). \]  
(6)

The hyperedge matrix \( D_e \) is defined as:
\[ D_e = \begin{bmatrix} \sigma(e_1) \\ \sigma(e_2) \\ \vdots \\ \sigma(e_n) \end{bmatrix}. \]  
(7)

The degree of a vertex is defined as:
\[ d(v) = \sum_{e \in E} w(e).h(v, e). \]  
(8)

The vertex degree matrix \( D_v \) is defined as:
\[ D_v = \begin{bmatrix} d(v_1) \\ d(v_2) \\ \vdots \\ d(v_n) \end{bmatrix}. \]  
(9)

3.2. The Construction Method of Hypergraph Model.
Fuzzy clustering is performed on the vertices of the hypergraph, and then the hyperedges are used to connect various vertices in turn. The number of hyper-edges of the hypergraph is the number of clustering categories, and a hyper-edge center point needs to be determined before using the incidence matrix. Figure 1 is a schematic diagram of a hypergraph model constructed based on the clustering method.

Because K-nearest neighbors use hyperedges to connect the vertices near \( K \) to each vertex, then the number of hyperedges in the hypergraph is the number of vertices. Before calculating the incidence matrix \( H \), the center point of the superedge needs to be fixed. For the superedge \( e_i \), the vertex \( V_i \) needs to be selected as the center point. Figure 2 is a hypergraph model diagram constructed based on the \( k \)-nearest neighbor method.

3.3. Hypergraph Sorting Algorithm. The NP problem is transformed into an actual numerical calculation optimization problem, and the loss function is defined as
\[ \Omega(f) = \frac{1}{2} \sum_{e \in E(u \in v, e)} \frac{w(e) h(u, e) h(v, e)}{\sigma(e)} \left( \frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2. \]  
(10)

Minimizing the loss function can ensure that the vertices in the same superedge have greater similarity, and the vertices in different superedges are less similar. According to formula (10), the following can be derived:
\[ \Omega(f) = \sum_{e \in E(u \in v, e)} \frac{w(e) h(u, e) h(v, e)}{\sigma(e)} \left( \frac{f(u) f(v)}{\sqrt{d(u) d(v)}} \right)^2. \]

If an initialization label vector \( y \) is given, the closer the score vector \( f \) to \( y \) is, the more accurate the calculation is. Therefore, the regular term is defined as:
\[ R(f) = \| f - y \|^2 = \sum_{u \in V} (f(u) - y(u))^2. \]  
(12)
The calculated $f$ will be solved by the optimal problem:
\[
\arg \min \int e^{R_i} [\Omega(f) + \mu R(f)]
\]
\[
= \arg \min \int e^{R_i} [f^T (I - \Theta)f + \mu \|f - y\|^2].
\]  
(13)

Solve by formula (13):
\[
f = (1 - \lambda) (I - \lambda \Theta)^{-1} y.
\]  
(14)

For formula (14), this formula is suitable for the use of the sorting calculation method.

3.4. Label Expansion Method. Research scholars sorted the vertices of the hypergraph by the random walk method, which made this method popularized on the hypergraph. The random walk process based on the weighted hypergraph is as follows: First, the starting superpoint is selected according to the superedge weight $w(e)$ Choose one that contains the current superpoint $P$. Among the superedges that have been selected, select the transition vertex $v$ according to the weight of the super point. Let $P$ be the transition probability matrix of a random walk. The calculation method is shown in formula (15):
\[
P(u,v) = \sum_{e \in E} w(e) \times \frac{h(u,e)}{\sum_{e \in E} w(e)} \times \frac{h(v,e)}{\sum_{e \in E} h(v,e)}.
\]  
(15)

If all nodes are visited, the probability distribution vector will always remain at 1 state. Only under the premise that Markov chains are non-periodic and irreducible can the above objectives be achieved. In order to meet these two conditions, this algorithm completes the random walk of all nodes through the PageRank algorithm under the injection of the mind transfer method. The data transfer in the previous article means that the superpoint located on any superedge may teleport to another superedge with a small probability. It is true that the two superpoints may not be connected, so it is impossible to transfer them directly. In this paper, this small probability is expressed by a damping factor, as shown in formula (16).
\[
v^{j+1} = (1 - \beta)P^T v^j + \frac{\beta e}{n}.
\]  
(16)

4. English News Text Recommendation Method Based on Tag Probability Correlation

4.1. Probabilistic Correlation Constructs Label Similarity Matrix. Looking at the set of user tags, tags and tags are not completely independent. There is a certain relationship between them. This potential relationship makes each tag different in importance to different users. In addition, the ambiguity of the label will make the label ambiguous in characterizing user characteristics. In part of the label set of user A, the label "apple" contains multiple meanings. It is necessary to clarify whether the label Apple is fruit or an electronic device. The calculation of the correlation between the probability, and then understand the user's tendency to use.

4.1.1. Tag Probability Correlation. From the overall observation, if any two tags are often marked by the user together, then these two tags have a great correlation. From local observations, if a label is marked by the user, the probability of another label being marked by the user is also very high, then it is inferred that the two labels have a strong co-occurrence relationship, as defined in formula (17).
\[
p(t_i|t_j) = p(t_j|t_i),
\]  
(17)

It can be seen from formula (17) that the conditional probability between tags is an asymmetric value. However, there is a symmetrical similar relationship between tags and tags. Therefore, it is necessary to improve the common relationship of tags, and the specific method adopted is shown in formula (18).
\[
\text{cor}(t_i, t_j) = p(t_i|t_j) \times p(t_j|t_i).
\]  
(18)

Combining formulas (17) and (18), the probability correlation between label $f$ and label $f$ is rewritten as
\[
\text{cor}(t_i, t_j) = \frac{p(t_i|t_j)^2}{p(t_i)} \times \frac{p(t_j)}{p(t_j)}.
\]  
(19)

4.1.2. Label Similarity Matrix. The label matrix is used to characterize users in the vector space model, but traditional algorithms can only reach the level of calculating the similarity between user vectors. Moreover, restricted by the extreme sparseness of user vectors, traditional methods cannot measure the similarity between tags well. Therefore, this paper obtains the probability correlation between tags by calculating the tag correlation phasor. Each tag in the tag set can be represented as a tag correlation vector, defined as formula (20).
\[
t_i = [\text{cor}(t_i, t_1), \text{cor}(t_i, t_2), \ldots, \text{cor}(t_i, t_n), \ldots, \text{cor}(t_i, t_K)].
\]  
(20)

Use the formula of cosine similarity to calculate the similarity between tags, as shown in formula (21).
\[
\text{sim}(t_i, t_j) = \frac{t_i \cdot t_j}{\|t_i\| \times \|t_j\|} = \sqrt{\frac{\sum_{i=1}^{K} \text{cor}(t_i, t_j) \text{cor}(t_j, t_i)}{\sqrt{\sum_{i=1}^{K} \text{cor}(t_i, t_j)^2} \sqrt{\sum_{i=1}^{K} \text{cor}(t_j, t_i)^2}}}
\]  
(21)

A similar label matrix $S$ is constructed according to the scheme of label weighting, and then $S$ is used to represent the cosine similarity between the two label phasors. The calculation formula is shown in formula (22).
\[
S_{ij} = \begin{cases} 
1, & i = j, \\
\text{sim}(t_i, t_j), & i \neq j.
\end{cases}
\]  
(22)
4.2. User Tag Weight. The labels that English news users label themselves are distributed as a whole, that is, a small number of representative labels are often labeled, while other personalized labels are usually rarely labeled, which leads to the label weight weighting scheme in the traditional label weight. The probability below is 1. In order to change the problem of repeated results caused by traditional calculation, a label weight weighting scheme is proposed: relevance weight. For English news readers who are tagged after browsing, there is a correlation between the tagged tags on the user. If a certain label of a user has a strong association with any other label, it indicates that the label has a strong identification degree for the user, and the definition is shown in formula (23).

$$\text{cow}(u_j, t_k) = \frac{\sum \text{cor}(t_k, t_j)}{|u_j|}. \quad (23)$$

Formula (23) only calculates the weight of the label to the user from the local features of the label. A comprehensive label weight must not only take into account the relationship between itself and other labels, but also consider the label to the user from the entire English news collection. The identity of is named inverse user frequency IuF. The specific idea is to use the ratio of the total number of users in the data set to the number of users adding a certain tag and take the logarithm,

$$\text{uF}(t_k) = \log_2\left(\frac{N}{\text{uf}(t_k)} + 1\right). \quad (24)$$

Combining the correlation weight and the value of $f$, the weight of the tag in the user $r$ is

$$w_{ik} = \text{cow}(u_i, t_k) \times \text{uF}(t_k). \quad (25)$$

Considering the probabilistic correlation between tags, a similarity matrix between the old and new tags can be formed, thereby giving users a new experience without creating a sense of alienation from the previous tags. The label matrix can not only remove the limitations of the original matrix, but also contains rich semantic information. Formula (26) is the updated user label matrix.

$$M' = M \times S = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1k} \\ w_{21} & w_{22} & \ldots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \ldots & w_{NK} \end{bmatrix} \begin{bmatrix} sim_{11} & sim_{12} & \ldots & sim_{1k} \\ sim_{21} & sim_{22} & \ldots & sim_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ sim_{k1} & sim_{k2} & \ldots & sim_{kk} \end{bmatrix}. \quad (26)$$

In order to better explain the updated matrix sparse problem, the decomposition matrix $M$ is as follows:

$$M' = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1k} \\ w_{21} & w_{22} & \ldots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \ldots & w_{NK} \end{bmatrix} + \begin{bmatrix} 0 & sim_{12} & \ldots & sim_{1k} \\ sim_{21} & 0 & \ldots & sim_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ sim_{k1} & sim_{k2} & \ldots & 0 \end{bmatrix}. \quad (27)$$

Define the ranking function of the recommendation algorithm:

$$f(u_i, d_p) = \frac{d_p \cdot u_i}{\|d_p\| \times \|u_i\|}. \quad (28)$$

4.3. Experimental Operating Environment. System physical operating environment is show in Table 1.

4.4. The Influence of Parameter Settings on Method Performance. The following will examine the influence of parameter ports and thresholds on the algorithm performance in the experiment. When testing the influence of one parameter value on the algorithm, keep the other parameter value unchanged. When weighting the superedge, adjust the ratio of the time factor and the popularity index in English news at any time. The calculation results show that the higher the value, the greater the impact of the time of publishing English news on the extraction of user interest,
and the smaller the value, indicating the number of comments and reposts in English news have a low impact on user interest. In this paper, different values of $a$ are set to compare the results of the recommended results at different ports, and the parameter is set to 0.5. When different values are taken, the performance is compared. Figure 3 shows the experimental results.

The size of the threshold determines the amount of English news recommended by the recommendation method to users, and different thresholds will have different recommended news numbers. In order to further study the influence of the threshold value on the experimental results, the threshold value is 0.8 for preliminary debugging, and then the threshold value is adjusted according to a certain ratio. The experimental results obtained by the test data set calculation method are shown in Figure 4.

It can be seen from Figure 4 that as the threshold increases, the accuracy of the algorithm gradually increases, while the recall rate of the algorithm gradually decreases.

4.5. Comparison of User Label Details and Performance before and after Label Expansion. Due to space limitations, the following is only a comparison of the front and back tags of 10 users as a display, as shown in Table 2.

It can be seen that the reason why these tags are frequently used is that these tags may appear in the user’s input items as recommended items, so customers may choose these tags according to their personal ideas, and the second is recommended tags. The universal applicability makes people more inclined to directly choose from the existing labels as closely as possible to their own labels. Comparison of different recommendation algorithms as shown in Table 3 (Table 3 is reproduced from Huifang et al. [18]).

As can be seen from the table, the LeALpc algorithm proposed in this paper is a higher level of accuracy than the CTR algorithm, UEMR algorithm and BPACMR algorithm from the content perspective, as well as the LPC algorithm, LC and ILCAUSR algorithm from the label perspective. Because these algorithms only consider the needs of users from an overall perspective, but cannot recommend good news based on the user’s personal experience, which leads to the problem of low hit rates in the recommendation list. However, the LeALpc algorithm relies on its unique operation mode to provide users with a better experience. It captures the attractiveness of the English news text and user tags to users. Therefore, in real life, the order of recommending news to users is often more important than the news itself. In other evaluation indicators, the LeALpc algorithm is significantly higher than the five algorithms except the ILcAusR algorithm, but it has no outstanding advantages compared to the ILcAusR algorithm. This is because the ILcAusR algorithm integrates the social relationships between users into the English news recommendation algorithm, which accurately caters to users’ interests. But this article has not considered it, which is the direction of this article’s future research.

4.6. Functional Test and Result Analysis for News Clustering.

In order to test the general effect of news clustering, some standards are proposed here, including the accuracy and recall rate of news, $F$ value, missed detection rate, false detection rate and normalized detection cost. Event discovery results as shown in Table 4.

In this experiment, change the value of $\tau$ from 0.2 to 0.8 in turn, adjust the step size to 0.1, and observe the changes of $P_{miss}$, $P_F$ and $(C_{det})_{norm}$ values of each part to determine the optimal $\tau$ value. The clustering algorithm uses the traditional hierarchy Clustering Algorithm, result as shown in Table 5.

This experiment carried out a comparative analysis experiment on the results of the whole network news data SogouCA obtained from 8 columns in the experimental data corpus of only SVM and LDA and the new model after the hybrid innovation of the two. Finally, it was calculated by the traditional hierarchical clustering algorithm. The final experimental results are shown in Figure 5:

It can be seen from the experimental results that the final $F$ value obtained by the two fusion models is more convincing than other models. The SVM and LDA models used in this article are proportionally weighted and the method of calculating model similarity is more convenient and accurate.

The following is an experiment on the proportion of the farthest distance between cluster news and the distance between cluster centers.

According to the data in Table 6, when the proportions of the distance between cluster centers and the farthest distance between clusters are 4/7 and 3/7 respectively, the clustering effect reaches an excellent level. Research on the calculation results found that the larger the proportion of the news distance between clusters, the clusters containing more data points are difficult to merge, so that similar news cannot be merged into the same cluster. If the distance between the cluster centers is large, the difference between the cluster and the cluster. The increase in the number of mergers easily leads to the occurrence of large clusters.

The following is a comparative experiment comparing the execution effect of traditional and improved hierarchical clustering algorithms. In operation, according to the
variance between the two algorithms, the accuracy of the experimental calculation data and the closeness of the calculation data sets are estimated, thereby Comparison results in better algorithms. Comparison of two clustering algorithms as show in Table 7.

4.7. Functional Test and Result Analysis for News Recommendation. This article uses the relevant data of the CCF big data competition “Analysis of User Browsing News and Personalized News Recommendation” as the experimental data of the recommendation algorithm. These data come from Caijing.com, and the calculated sample is
selected as all the news browsing information of 5,000 registered users of the website randomly selected in the second half of 2020. The calculation file has 6 basic attributes, including user number, browsing time, news number, news title, news publication time and its detailed content. 95% of the experimental data is used as the training dataset, and 5% as the test data set.

In the course of this experiment, the recommendation list is generated using the TopN strategy, that is, the top N news events with the highest ratings are selected and recommended to users. This personalized recommendation does not take the topic model as the key point, but takes the personalized news discovery and recommendation that can capture the user experience on demand as the focus of the evaluation. Therefore, it can be seen that the personalized news recommendation technology used in this article will have a greater advantage in solving the user’s cosine similarity.

This experiment adjusts the $\rho$ value from 0.6 to 1.0 in turn, and compares and analyzes the accuracy of the algorithm recommended in this paper, and obtains the best $\rho$ value. The experimental results are shown in Figure 6.

The following is to verify the influence of the percentage threshold on the recommendation result, and calculate the recall rate, accuracy rate and $F$ value generated under the use of the recommendation algorithm in this article. The specific representation is shown in Figure 7.

The following experiments are based on the impact of different interest modeling methods on the recommendation results, considering the three basic elements of browsing event time, popularity, and event update frequency to construct a new model. This experiment will use quantitative

### Table 3: Comparison of different recommendation algorithms.

| Name   | $L = 5$ |  | $L = 10$ |  | $L = 15$ |  | $L = 20$ |  |
|--------|---------|---|----------|---|----------|---|----------|---|
|        | $p$     | AP|          |   | $p$      | AP|          |   |
| LC     | 0.315   | 0.556 | 0.358     | 0.425 | 0.358 | 0.548 | 0.286 | 0.438 |
| LPC    | 0.306   | 0.517 | 0.392     | 0.486 | 0.409 | 0.486 | 0.245 | 0.451 |
| ILCAUSR | 0.426  | 0.590 | 0.455     | 0.542 | 0.456 | 0.592 | 0.263 | 0.538 |
| CTR    | 0.385   | 0.482 | 0.412     | 0.563 | 0.475 | 0.575 | 0.275 | 0.514 |
| UEMR   | 0.352   | 0.476 | 0.453     | 0.486 | 0.446 | 0.565 | 0.285 | 0.495 |
| BPACMR | 0.342   | 0.586 | 0.393     | 0.514 | 0.450 | 0.541 | 0.263 | 0.517 |
| LeALpc | 0.338   | 0.581 | 0.445     | 0.599 | 0.463 | 0.628 | 0.272 | 0.553 |

### Table 4: Event discovery results.

| Category                  | Number of related documents | Number of irrelevant documents |
|---------------------------|-----------------------------|--------------------------------|
| The number of documents detected | A                          | B                              |
| Number of documents not detected | C                        | D                              |

### Table 5: $P_{\text{miss}}$, $P_{\text{false}}$, and $(C_{\text{det}})_{\text{norm}}$ changes with $\tau$.

| $\tau$ | $P_{\text{miss}}$ | $P_{\text{false}}$ | $(C_{\text{det}})_{\text{norm}}$ |
|--------|------------------|---------------------|----------------------------------|
| 0.2    | 0.0806           | 0.0763              | 0.8576                           |
| 0.3    | 0.0032           | 0.0026              | 0.7112                           |
| 0.4    | 0.0039           | 0.0033              | 0.0806                           |
| 0.5    | 0.0766           | 0.0723              | 0.0024                           |
| 0.6    | 0.0766           | 0.0723              | 0.0024                           |
| 0.7    | 0.0766           | 0.0723              | 0.0024                           |
| 0.8    | 0.0766           | 0.0723              | 0.0024                           |

### Table 6: Comparison of cluster center distance and proportion.

| Weight of distance between cluster centers | The weight of the farthest news distance | Calculation accuracy |
|-------------------------------------------|----------------------------------------|----------------------|
| 0                                         | 1                                      | 68.87%               |
| 1/7                                       | 6/7                                    | 75.08%               |
| 2/7                                       | 5/7                                    | 81.06%               |
| 3/7                                       | 4/7                                    | 84.68%               |
| 4/7                                       | 3/7                                    | 87.96%               |
| 5/7                                       | 2/7                                    | 72.39%               |
| 6/7                                       | 1/7                                    | 66.16%               |
| 1                                         | 0                                      | 61.67%               |

---

**Figure 5:** Comparison of each model based on $F$ value.

---

**Figure 6:** Comparison of cluster center distance and proportion.

---

**Figure 7:** Comparison of event discovery results.
analysis and chart methods to compare the differences between traditional algorithms and improved algorithms, and then highlight the advantages of the methods described in this article.

As can be seen from Figure 8, the modeling method presented in this article can completely overwhelm the traditional modeling method with huge advantages at all times.

As can be seen from Figure 9, the modeling method of this article is significantly better than the traditional modeling method.

As can be seen from the above Figure 10, the modeling method in this article has greatly improved compared with traditional modeling methods in terms of recall rate, accuracy rate, or F value. Through the above series of experiments and research, it is completely possible to prove that the improved modeling method has great technical advantages.

The following is an experiment that compares the accuracy, recall, and F value of the three algorithms in this article, the content-based recommendation algorithm, and the collaborative filtering recommendation algorithm.
The comparison results are shown in Figure 11 after calculation:

According to Figure 11, the recommendation algorithm in this paper has more use value than algorithms with other calculation methods in terms of recall rate, accuracy rate or F value.

5. Conclusion

This paper introduces the hypergraph model into the field of English news text summarization, and proposes an English news text summarization algorithm based on hypergraph sorting and a summary algorithm based on hypergraph random walk. Both algorithms capture the complex relationship between English news texts through the hypergraph model. The HGRVS algorithm uses the idea of hypergraph sorting to classify the English news texts, and the English news texts obtained from the candidates are obtained after a certain calculation. The static English text that I want to generate, but RWH uses a hypergraph with random walk characteristics to sort the English text in order, and then form a static English text summary based on the order. The information is extremely expanded, and it is difficult for Internet users to use only “portal sites and search engines” to find their own suitable news world. However, the technology recommended in this article combines news discovery and personalized recommendation, which can not only help users get news quickly, but also users do not need to worry about and waste time searching for similar news, thus providing users with a better reading experience. The new news recommendation technology proposed in this paper is different from the single recommendation method that only appeared in the previous news recommendation methods. This calculation method first uses the hierarchical clustering algorithm based on the fusion model to collect news events, and then integrates the user’s usage and the multiplicity of the events, and calculates the user’s interest model, and finally will use the hybrid recommendation algorithm. After the calculation, the located news events are sent to the user. After the user reads, the hierarchical clustering of the news interval calculation method is applied to the reading results to calculate more similar articles suitable for the user. This method uses TF-IDF as the basis for similarity calculation, and then expands the calculation of the two topic models, SVM and LDA, with a certain proportion of weighted sum. From the large number of experimental results in this article, comparing the previous unimproved algorithms with the improved algorithm, we can understand that the recommendation algorithm provided in this article has great advantages over other algorithms. To solve the problem of finding similar articles for users, this paper adopts a method for calculating the distance between clusters that can be used for hierarchical clustering. This method adjusts the proportion of the distance between the cluster center and the farthest distance between the clusters in the formula, without affecting the calculation accuracy, and has a great advantage over traditional calculation methods that cannot solve the problem of large clusters.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Zarzour, F. Mzaouzou, M. Soltani, and C. Chemam, “An improved collaborative filtering recommendation algorithm for big data,” Computational Intelligence and Its Applications, vol. 12, no. 1, pp. 660–668, 2018.

[2] L. O. Colombo-Mendoza, R. Valencia-García, R. Colomopalacios, and G. Alor-Hernández, “A knowledge-based multi-criteria collaborative filtering approach for discovering services in mobile cloud computing platforms,” Journal of Intelligent Information Systems, vol. 54, pp. 1–25, 2018.

[3] C. Lu, “Recommendation algorithm based on collaborative filtering under social network environment,” Bio-Technology: An Indian Journal, vol. 10, no. 13, pp. 7087–7092, 2014.

[4] Z. Chen, Z. Ning, Q. Xiong, and M. S. Obaidat, “A collaborative filtering recommendation-based scheme for WLANs with differentiated access service,” IEEE Systems Journal, vol. 12, no. 1, pp. 1004–1014, 2018.

[5] H. Zarzour, Z. A. Sharif, M. A. Ayyoub, and Y. Jararweh, “A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques,” in Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS), pp. 102–106, Lille, France, April 2018.

[6] Q.-J. Tan, H. Wu, C. Wang, and Q. Guo, “A personalized hybrid recommendation strategy based on user behaviors and its application,” in Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics, pp. 181–186, SPAC, Shenzhen, China, December 2017.

[7] J. Stan, F. Muhlenbach, and C. Largeron, “Recommender systems using social network analysis: challenges and future trends,” in Encyclopedia of Social Network Analysis and Mining, pp. 1–22, Springer, New York, NY, USA, 2014.

[8] K. G. Saranya and G. Sudha Sadasivam, “Personalized news article recommendation with novelty using collaborative filtering based rough set theory,” Mobile Networks and Applications, vol. 22, no. 4, pp. 719–729, 2017.

[9] A. M. Rashid, K. L. Shyong, and K. L. George, “A highly scalable hybrid model & memory-based CF algorithm,” in Computational Intelligence and Neuroscience.
Proceeding of the WEBKDD of ACM, Macau, China, June 2014.

[10] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[11] J. M. Schleicher, M. Vogler, S. Dustdar, and C. Inzinger, "Enabling a Smart City application ecosystem: requirements and architectural aspects," *IEEE Internet Computing*, vol. 20, no. 2, pp. 58–65, 2016.

[12] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy protection for preventing data over-collection in Smart City," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1339–1350, 2016.

[13] T. Roh, K. J. Oh, and I. Han, "The collaborative filtering recommendation based on SOM cluster-indexing CBR," *Expert Systems with Applications*, vol. 25, no. 3, pp. 413–423, 2003.

[14] M. Hasan and B. Starly, "Decentralized cloud manufacturing-as-a-service (CMaaS) platform architecture with configurable digital assets," *Journal of Manufacturing Systems*, vol. 56, no. 2, pp. 157–174, 2020.

[15] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[16] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-based recommendation system using bayesian user’s preference model in mobile devices," in *Ubiquitous Intelligence and Computing*, vol. 4611, pp. 1130–1139, Springer, Berlin, Germany, 2007.

[17] C. Duanbing, Y. Fu, and M. S. Shang, "An efficient algorithm for overlapping community detection in complex," *Global Congress on Intelligent Systems*, vol. 10, pp. 244–247, 2009.

[18] M. Huifang, Z. Di, Z. Weizhong, and S. Zhongzh, "Weibo recommendation method based on hypergraph random walk label augmentation," *Journal of Software*, vol. 30, no. 11, pp. 3397–3412, 2019.