INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 19, Issue: 5, Month: October, Year: 2024
Article Number: 6658, https://doi.org/10.15837/ijccc.2024.5.6658

CCC Publications

AGORA UNIVERSITY PRESS

A Topic Recommendation Control Method Based on Topic Relevancy and R-tree Index

Jing Yu, Zhixing Lu, Xianghua Li, Bin Wu, Shunli Zhang, Zongmin Cui

Jing Yu
School of Management
Jiujiang University, China
Jiujiang, Jiangxi 332005, China

Zhixing Lu
1. Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, Malaysia
Serdang, Selangor 43400, Malaysia
2. School of Computer Science and Technology
Hainan University, China
Haikou, Hainan 570228, China

Xianghua Li*
School of Humanities
Anqing Normal University, China
Anqing, Anhui 246002, China
*Corresponding author: lixianghua136@aqnu.edu.cn

Bin Wu
School of Computer and Big Data Science
Jiujiang University, China
Jiujiang, Jiangxi 332005, China

Shunli Zhang
School of Computer and Big Data Science
Jiujiang University, China
Jiujiang, Jiangxi 332005, China

Zongmin Cui
School of Computer and Big Data Science
Jiujiang University, China
Jiujiang, Jiangxi 332005, China

Abstract

Topic recommendation control aims to suggest relevant topics to users based on their preferences and regional trends. However, existing methods often lack effective measures to evaluate topic-user relevancy and require comparing large amounts of regional information, leading to low accuracy and efficiency. Therefore, we propose a Topic Recommendation Control method based on topic Relevancy and R-tree index (named as TRCRR) to address these limitations. TRCRR introduces a novel personalized topic relevancy metric that quantifies the relevancy between topics and user preferences. To improve efficiency, an R-tree topic index is constructed to organize topics across different regions hierarchically. Experiments on a real-world dataset show that TRCRR achieves better recommendation accuracy and efficiency compared to several baseline methods. The proposed approach offers a promising solution for personalized and region-aware topic recommendation.

Keywords: topic recommendation, recommendation control, topic relevancy, R-tree index, regional communication.

1 Introduction

With the popularity of various social media and social software, user access records and natural language generate large amounts of data on certain topics [1]. Topic data with region tags [2, 3] contains a lot of valuable information that can be used in a variety of applications such as intelligent transportation [4, 5], intelligent healthcare [6, 7], embedded processor [8], language processing [9], recommendation systems [10], image recognition [11], etc. Allowing users to quickly understand the main content or current trends in topic data, the topic recommendation control method analyzes topic data to identify potential user needs, which has important application value [12, 13]. Therefore, controlling recommendations through data computation and communication has become a widely valued and concerned subject [14].

Topic recommendation control first utilizes specific algorithms and structures to mine potential and valuable user-preference topics in region data [15]. Then it determines personalized topics that users may need based on these preference topics and topic recommendation control algorithms [16]. The topics selected by the topic recommendation control algorithm often represent some mainstream trends within a certain region. Based on the selection, businesses can timely adjust their sales strategies, or researchers can determine the direction of future research according to these mainstream trends [17]. However, with the further increase in the amount of regional topic data, the user's preferred topics are becoming increasingly complex and diverse.

For example, a tourist is traveling in Chengdu city. When searching for nearby cuisine, the system may recommend some of the highest-rated restaurants in the region to the tourist. But the regional characteristic of Chengdu may be that the food is spicy, while the tourist likes to eat sweets. The difference between this preference and regional characteristics leads to tourists not liking the things recommended by the system. This results in low-quality personalized topic recommendations and insufficient recommendation services. High-quality topics have great reference value and can provide a lot of convenience for relevant personnel. However, low-quality topics may not only mislead businesses or researchers, but also greatly waste people's time and energy, and even economic costs, causing huge losses [18].

To enhance the quality of recommended topics, researchers construct personalized topic recommendation algorithms by utilizing regional data. There have been many related studies [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29], but there are two limitations with the existing regional topic recommendation control methods as follows. (1) There is a lack of a measure to evaluate the relevancy between candidate topics and user-preference topics. Lack of this relevancy may lead to low recommendation accuracy for recommending user-interesting topics to users. (2) Different regions generate a large amount of regional topic information. The information needs to be compared every time to control the topic recommendations that are of interest to users. The frequent comparison may lead to low recommendation efficiency. Therefore, we propose a Topic Recommendation Control method based on topic Relevancy and R-tree index (named as TRCRR) to address the above limitations. Our contributions are shown as follows.

- (1) We provide the personalized topic relevancy list as a measured approach to evaluate the relevancy between candidate topics and user-preference topics for enhancing the recommendation accuracy.
- (2) We design an R-tree topic index to reduce the aggregation size of topic lists in different regions. The index accelerates the topic recommendation control for increasing recommendation efficiency.
- (3) Extensive experiments show that our method TRCRR has better recommendation accuracy and efficiency than several baseline methods.

The structure of this paper is organized as follows. Section 2 provides the related works. Section 3 presents the preliminary. Section 4 designs the topic recommendation control. Section 5 analyzes the experiments and Section 6 concludes this paper.

2 Related Works

Spatial topic refers to topics in a physical space, which is a relatively large area. Location topic refers to topics in a geographical location, which is a smaller area. Therefore, the regional topic recommendation control could be roughly divided into the following two types: spatial topic recommendation control and location topic recommendation control.

2.1 Spatial topic recommendation control

Chen et al. [19] correct the loss of user topic expectations and reduce the optimization and refinement problem to a general linear programming problem. The goal is to optimize the orientation perception of keywords in recommendations. Yin et al. [20] propose a spatially aware hierarchical collaborative deep learning model. The model jointly performs deep representation learning from heterogeneous features and hierarchical additive representation learning from spatially perceived individual preferences. Zhou et al. [21] provide a topic-enhanced memory network, which is a deep architecture that integrates topic models and memory networks by leveraging the global structure of latent patterns and the advantages of nonlinear features based on local neighborhoods. Li et al. [22] design a spatio-temporal topic model for cold start event recommendations that captures user interest over time. A spatio-temporal embedded subject model is proposed in [23] to solve the recommendation problem of remote sensing images. The model constructs the topic model to make full use of the spatial and temporal continuity characteristics and improve the training efficiency of the recommendation model. Liu et al. [24] present a distributed spatiotemporal data processing system, named as ST4ML, to support scalable machine learning-oriented applications.

The different regions where topics are generated have a large amount of regional topic information in the spatial topic recommendation control. The massive information needs to be compared every time to control the topic recommendations that are of interest to users, which leads to low recommendation efficiency. Thus, we propose a R-tree topic index to reduce the aggregation size of topic lists in different regions to accelerate topic recommendation control.

2.2 Location topic recommendation control

Zhao et al. [25] propose a POI (Point of Interest) group recommendation method using an extreme learning machine, which treats POI group recommendation as a binary classification problem. A POI static feature extraction method based on symmetric matrix factorization is designed to capture location and POI category features in [26]. Canturk et al. [27] provide an undirected graph model. In the model, the recommended score for location is the result of performing random walks on trust-enhanced LBSN (Location-Based Social Networks) subgraphs. Gao et al. [28] introduce a location-aware information fusion model with a dual granularity human mobility learning module. Lv et al. [29] believe that location affects news recommendations based on region rather than latitude and longitude levels.

However, location topic recommendation control methods lack a measure to evaluate the relevancy between candidate topics and user-preference topics for recommending user-interesting topics to users. Therefore, we provide the measure approach to enhance the recommendation accuracy.

3 Preliminaries

Our topic recommendation control framework is shown in Figure 1. The framework is decentralized with two kinds of devices. The edge devices include smart sensors, programmable logic controllers, edge smart routers, etc. The edge devices are responsible for generating data and performing lightweight operations on the data. The edge controls lightweight and fast topic recommendations by communicating with each other and collaborating with the cloud. The cloud is responsible for the storage and calculation of massive data, and provides high-quality recommendation control services to the edge and users. In a word, the recommendation effect is controlled through cloud-edge collaborative communication. The main symbols and corresponding explanations used for TRCRR are given in Table 1.

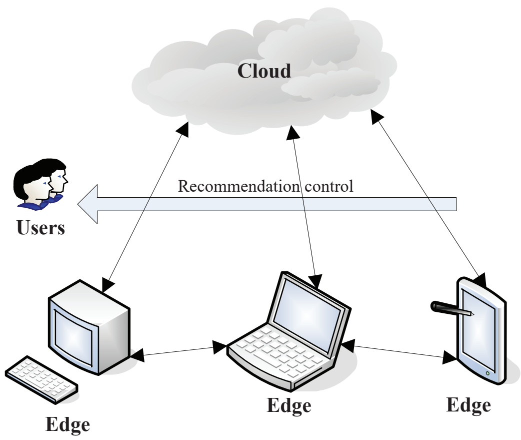

Figure 1: TRCRR framework.

Table 1: The main symbols for TRCRR

Symbol	Explanation
B	A set of objects in the system
b	An object in B
w	A topic in B
R	A set of regions in the system
r	A region in R
m_i	The number of occurrences of w in objects whose type is t_i
U	A set of users in the system
u	A user in U
w_u	User u's preferred topic
r_u	User u's interested region
k_u	The number of topics that user u may be interested in
Rel_{w_u,w_i}	Topic relevancy between w_u and w_i
$freq(p_w^r, x \times \|B\|)$	The frequency of topic w
$\theta(p_w^r, x \times \|B\|)$	The sum of frequencies ranked as p_w^r
$\delta(p_w^r, x \times \|B\|)$	Threshold
$\beta(p_w^r, x \times \|B\|)$	Intercepted length

After removing "StopWord" [30], the set of objects in the framework is represented as $B = \{b_1, b_2, b_3, \ldots, b_{n-1}, b_n\}$. An object b may be composed of a series of topics w_i, where $i \in R^+$. That is $b = \{w_1, w_2, w_3, \ldots, w_{n-1}, w_n\}$, where $w_i \notin$ "StopWord". Meanwhile, there is a type attribute for each object b and topic w defined as follows.

Definition 1 (Object type). *$\forall b \in B$, if there are the most topics with type t in object b, then the object type of b is represented as t, i.e. $type(b) = t$.*

Table 2: The corresponding topic information in entire region R

Object	Topic	Region
b_1	$[w_1, w_2, w_1, w_3, w_4, w_2, w_3, w_1]$	r_4
b_2	$[w_1, w_6, w_7, w_6, w_4]$	r_4
b_3	$[w_1, w_6, w_1, w_3, w_4, w_2, w_6, w_1]$	r_5
b_4	$[w_5, w_5, w_3, w_5]$	r_6
b_5	$[w_2, w_4, w_4, w_7, w_3]$	r_6
b_6	$[w_1, w_5, w_1, w_5, w_3, w_5]$	r_6
b_7	$[w_1, w_4, w_5, w_7, w_1, w_2, w_4, w_1]$	r_7
b_8	$[w_1, w_5, w_5, w_3, w_5, w_2, w_3, w_5]$	r_7
b_9	$[w_6, w_6, w_1, w_3, w_6, w_3, w_1]$	r_7
b_{10}	$[w_1, w_7, w_2, w_7, w_4, w_7, w_7, w_1]$	r_7

Definition 2 (Topic type)**.** *$\forall w \in b$, if topic w most frequently appears in an object with type t, the type of topic w is represented as t, i.e. type$(w) = t$.*

In Definitions 1 and 2, it is assumed that the two types both come from the same type set. In fact, these two types would iteratively change each other according to the update of regional data.

For each topic w , there are two attributes $< G, M >$, where G denotes the set of regional information and M denotes the set of number information. $G = [t, r, a, b]$, where t denotes the type of object b who owns w , r denotes the region of b and a denotes the number of occurrences of w in object b. $M = [m_1, m_2, m_3, \ldots, m_{n-1}, m_n]$. $\forall m_i \in M, m_i$ denotes the number of occurrences of w in objects whose type is t_i. Let $S_i = \{a|[t, r, a, b] = G, t = t_i\}$, then m_i can be calculated as Formula 1.

$$m_i = \sum_{i=1}^{|S_i|} a_i \tag{1}$$

Figure 2 shows a running example for ease of understanding the method presented in this paper. Table 2 shows the topic information corresponding to Figure 2. Based on the above definitions, there are 3 w_5 in b_4 and 3 w_5 in b_6. The types of b_4 and b_6 are both t_3. Thus, $m^{r_6}_{3-w_5} = 3 + 3 = 6$. Consequently, $M^{r_6}_{w_5} = [0, 0, 6, 0, 0]$. In the same way, r_6's and r_7's attributes corresponding to Table 2 are shown in Table 3.

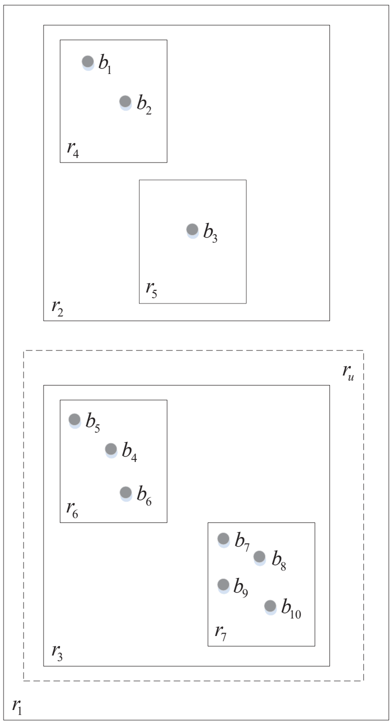

Figure 2: A running example for topic recommendation control.

Table 3: The corresponding attributes in r_6 and r_7

Topic	G	M
	r_6	
w_5	$[t_3, r_6, 3, b_4], [t_3, r_6, 3, b_6]$	$[0, 0, 6, 0, 0]$
w_3	$[t_3, r_6, 1, b_4], [t_2, r_6, 1, b_5], [t_3, r_6, 1, b_6]$	$[0, 1, 2, 0, 0]$
w_2	$[t_2, r_6, 1, b_5]$	$[0, 1, 0, 0, 0]$
w_4	$[t_2, r_6, 2, b_5]$	$[0, 2, 0, 0, 0]$
w_7	$[t_2, r_6, 1, b_5]$	$[0, 1, 0, 0, 0]$
w_1	$[t_3, r_6, 2, b_6]$	$[0, 0, 2, 0, 0]$
	r_7	
w_1	$[t_1, r_7, 3, b_7], [t_3, r_7, 1, b_8], [t_4, r_7, 2, b_9], [t_5, r_7, 2, b_{10}]$	$[3, 0, 1, 2, 2]$
w_4	$[t_1, r_7, 2, b_7], [t_5, r_7, 1, b_{10}]$	$[2, 0, 0, 0, 1]$
w_5	$[t_1, r_7, 1, b_7], [t_3, r_7, 4, b_8]$	$[1, 0, 4, 0, 0]$
w_7	$[t_1, r_7, 1, b_7], [t_5, r_7, 4, b_{10}]$	$[1, 0, 0, 0, 4]$
w_2	$[t_1, r_7, 1, b_7], [t_3, r_7, 1, b_8], [t_5, r_7, 1, b_{10}]$	$[1, 0, 1, 0, 1]$
w_3	$[t_3, r_7, 2, b_8], [t_4, r_7, 2, b_9]$	$[0, 0, 2, 2, 0]$
w_6	$[t_4, r_7, 3, b_9]$	$[0, 0, 0, 3, 0]$

Table 4: The corresponding relevancy list between w_u and topics in r_6 and r_7

Relevancy list in r_6		Relevancy list in r_7	
w_1	1	w_6	3
w_2	2	w_2	5
w_7	2	w_5	5
w_4	3	w_1	6
w_5	5	w_4	7
		w_7	9

Definition 3 (Topic recommendation control)**.** *Let $< w_u, r_u, k_u >$ be user u's preferences information, where w_u denotes user u's preferred topic, r_u denotes user u's interested region, and k_u represents the number of topics that user u may be interested in. Topic recommendation control is the process of filtering out approximate k_u personalized topics that are most relevant to w_u in region r_u. The filtered topics are from B for being recommended to user u.*

In Figure 2, $r_u = r_3$, $r_6 \subset r_3$, and $r_7 \subset r_3$. Therefore, we need to only focus on r_6 and r_7 whose attributes are shown in Table 3. The topic recommendation control in Figure 2 is to identify some topics (from objects shown in Table 3) that users are most likely to be interested in.

4 Topic Recommendation Control

In the topic recommendation control, the source of the topic dataset is not a single one, but is composed of multiple similar social platform data, literature libraries, or other databases. This processing approach achieves cross-platform integration of topic data and solves the problem of low-quality recommended topics due to the reliance on a single topic dataset.

Based on the above definitions, to compare the similarity between two topics, we use subtraction at the same position in the information M of two topics to evaluate the relevancy between the two topics. Therefore, our personalized topic relevancy between w_u and w_i is calculated as Formula 2, where m_j^u denotes $j-th$ number in M of user u. The relevancy with the smallest value indicates that w_i has the highest similarity with w_u.

$$Rel_{w_u,w_i} = \sum_{j=1}^{|M|} |m_j^u - m_j^i| \tag{2}$$

Taking Table 3 as an example, if user u's preferred topic is w_3 (i.e. $w_u = w_3$), $Rel^{r_6}_{w_3,w_5} = |1-0| + |2-6| = 5$. In the same way, the corresponding relevancies between w_u and topics in r_6 and r_7 are shown in Table 4. Obviously, w_1 has the highest similarity with w_u (i.e. w_3).

Next, we need to determine how many and which topics should be returned to the user. Assuming that each object b has an average of x topics, then the total number of topics in the topic dataset is $x \times |B|$. Zipf's law [31] states that the frequency of a topic is inversely proportional to its ranking in the frequency list. Therefore, the frequency of topic w is calculated as Formula 3, where p_w^r denotes

Algorithm 1 $R - tree - Construction$

Input: R

Output: $R - tree, H$

1: $R = \{r_1, r_2, r_3, \ldots, r_{n-1}, r_n\}$
2: $h = 1$
3: $R - tree = \emptyset$
4: **while** $|R| \neq 0$ **do**
5:     **if** $\exists R^h \subset R \wedge \forall r_i \in R^h \wedge \nexists r_j \in R$ with $r_j \subset r_i$ **then**
6:         $R = R \diagup R^h$
7:         $h + +$
8:     **end if**
9: **end while**
10: $h - -$
11: $H = h - 1$
12: $R - tree = R - tree \cup R^h$
13: **while** $h \neq 1$ **do**
14:     $k = h$
15:     $h - -$
16:     $R - tree = R - tree \cup R^h$
17:     **for** $r_i \in R^h$ **do**
18:         **if** $r_j \in R^k \wedge r_i \subset r_j$ **then**
19:             connect directed line from r_j to r_i for $R - tree$
20:         **end if**
21:     **end for**
22: **end while**
23: **return** $(R - tree, H)$

w's ranking in the relevancy list in region r. For example, w_1 is ranked first in the relevancy list in r_6. Thus, $p_{w_1}^{r_6} = 1$ in Table 4.

$$freq(p_w^r, x \times |B|) = \frac{x \times |B|}{p_w^r} \tag{3}$$

In the specific implementation process, it is necessary to first divide the entire region R into multiple small regions, and then use the inclusion relationship between regions to establish an $R - tree$ topic index. The construction process is shown in Algorithm 1, where the entire region R is the input and the constructed $R - tree$ and layer height H are the output.

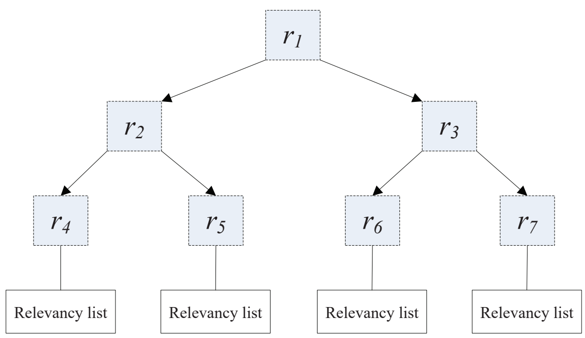

Figure 3: The $R - tree$ corresponding to Figure 2.

First, Algorithm 1 finds leaf nodes and places them into R^1 (Step 5). Second, it removes R^1 from R (Step 6). Similarly, it sequentially obtains the nodes of each layer (Steps 4-9). Third, it calculates layer height H (Step 11). H excludes the root node. Fourth, it puts the nodes of each layer into $R - tree$ (Step 16). Finally, starting from the root node, it gradually connects nodes based on the

Algorithm 2 TRCRR

Input: R, U
Output: T_U
$T_U = \emptyset$
$B = \{b_1, b_2, b_3, \ldots, b_{n-1}, b_n\}$
for $b_i \in B$ **do**
  for $w \in b_i$ **do**
    Construct $< G, M >$ for w
  end for
end for
$(R - tree, H) = R - tree - Construction(R)$
for $u \in U$ **do**
  $T_u = \emptyset$
  for $r \in R \wedge w_i \in r \wedge w_u \neq w_i$ **do**
    $Rel_{w_u,w_i} = \sum_{j=1}^{|M|} |m_j^u - m_j^i|$
  end for
  $\delta(p_w^r, x \times |B|) = \frac{R^u \times x \times |B_u|)}{k_u}$
  $\theta(p_w^r, x \times |B|) = 0$
  for $p_w^r \in [1, +\infty]$ **do**
    for $i \in [1, H]$ **do**
      $\theta(p_w^r, x \times |B|) += \frac{R^{i^u} \times x \times |B|}{p_w^r \times |R_i|}$
    end for
    if $\theta(p_w^r, x \times |B|) \leqslant \delta(p_w^r, x \times |B|)$ **then**
      $\beta(p_w^r, x \times |B|) = p_w^r$
      Exit
    end if
  end for
  for $r \in R$ **do**
    Select the top $\beta(p_w^r, x \times |B|)$ topics from r and put them into T_u
  end for
  $T_U = T_U \cup T_u$
end for
return T_U

inclusion relationship to form $R - tree$ (Steps 17-21). The time complexity of Algorithm 1 is $O(|R|)$.

Taking Figure 2 as an example, $R^1 = \{r_4, r_5, r_6, r_7\}$, $R^2 = \{r_2, r_3\}$, and $R^3 = \{r_1\}$. Meanwhile, r_3 includes r_6 and r_7, so r_6 and r_7 are leaf nodes of r_3. In the same way, the constructed $R - tree$ is shown in Figure 3. Meanwhile, layer height $H = 3 - 1 = 2$.

In the $R - tree$, there are a total of H layers from the leaf nodes up, then the $i - th$ layer has $|R^i|$ nodes. In the $i - th$ layer, there are R^{i^u} nodes that user u is interested in. Meanwhile, R^u denotes all nodes that user u is interested in excluding the root node. For example, Layer 1 R^1 has 4 nodes in Figure 3, thus $|R^1| = 4$. Layer 1 has 2 nodes that user u is interested in: r_6 and r_7, so $R^{1^u} = 2$. In the same way, $R^u = 3$ (i.e. r_3, r_6 and r_7 excluding root node r_1).

Based on the above definitions, the sum of frequencies ranked as p_w^r is calculated as Formula 4.

$$\theta(p_w^r, x \times |B|) = \sum_{i=1}^{H} \frac{R^{i^u} \times x \times |B|}{p_w^r \times |R^i|} \tag{4}$$

Then the threshold δ is calculated as Formula 5, where k_u is the number of topics that user u may be interested in based on the retrieved result (shown in Definition 3). Meanwhile, $|B_u|$ denotes the number of objects in user u's interested region. For example in Table 2, user u's interested region $r_u = r_3$. As $r_6 \subset r_3$ and $r_7 \subset r_3$, there are 7 objects in r_3: $\{b_4, b_5, b_6, b_7, b_8, b_9, b_{10}\}$. Thus, $|B_u| = 7$.

Table 5: Experimental parameters

	Effective topics	**Noneffective topics**
Retrieved	Effective topics are judged to be effective (EE)	Noneffective topics are judged to be effective (NE)
Not retrieved	Effective topics are judged to be noneffective (EN)	Noneffective topics are judged to be noneffective (NN)

Table 6: Evaluation criteria

Effectiveness	$\frac{EE+NN}{EE+NE+EN+NN}$
Precision	$C = \frac{EE}{EE+NE}$
Recall	$L = \frac{EE}{EE+EN}$
F1-measure	$\frac{2\times C\times L}{C+L}$

$$\delta(p_w^r, x \times |B|) = \frac{R^u \times x \times |B_u|}{k_u} \tag{5}$$

Finally, intercepted length β is calculated as Formula 6.

$$\beta(p_w^r, x \times |B|) =_{p_w^r}^{min} \{\theta(p_w^r, x \times |B|) \le \delta(p_w^r, x \times |B|)\} \tag{6}$$

Based on the above formulas, Algorithm 2 shows our topic recommendation control method TRCRR. It takes a set of regions R and a set of users U as input and a set of topics T_U returned to users U as output.

First, Algorithm 2 generates two attributes $< G, M >$ for each topic w (Steps 2-7). Second, it calls Algorithm 1 to construct the $R - tree$ corresponding to R (Step 8). Third, it calculates the relevancy list (Steps 11-13). Fourth, it calculates the threshold δ (Step 14). Fifth, it gets the sum of frequencies ranked as p_w^r (Steps 15-19). Sixth, it computes the intercepted length $\beta(p_w^r, x \times |B|)$ (Steps 20-23). Finally, it selects the top $\beta(p_w^r, x \times |B|)$ topics from each region r and put them into T_U (Steps 25-28). The time complexity of Algorithm 2 is $O(|U|\times|R|)$.

For example, $x = 6.7$ and $|B|= 10$ in Table 2. Thus, $\theta(p_w^r, x \times |B|) = \frac{2\times6.7\times10}{p_w^r\times4} + \frac{1\times6.7\times10}{p_w^r\times2} = \frac{67}{p_w^r}$. If $k_u = 5, \delta(p_w^r, x \times |B|) = \frac{3\times6.7\times7}{5} = 28.1$. Obviously, when $p_w^r = 3$, $\theta(p_w^r, x \times |B|) = 22.3 \le 28.1$. Therefore, $\beta(p_w^r, x \times |B|) = 3$. Algorithm 2 selects the top 3 topics from r_6 and r_7 based on the relevancy list shown in Table 4. Finally, $T_u = \{w_1, w_2, w_7, w_6, w_5\}$.

5 Experimental Analysis

5.1 Experiment setup

To verify the validity and practicality of our method TRCRR, we conduct extensive experiments using a real topic dataset, which includes five types. Each type has approximately 1000 objects containing titles, totaling approximately 500000 words (including duplicate and miswritten words). The data source is from http://cul.news.sina.com.cn/ .

This paper uses the top 6000 topics in the relevancy list as experimentally discoverable topics. The effective topics are those that appear more than 1 time in the relevant object titles of the topic dataset. The experimental environment for this experiment is a portable laptop. The configuration of the laptop is as follows. The operating system is Windows 10, the processor is Intel(R) Core(TM) i7-8750H CPU @2.20GHZ 2.21GHZ and the editor is Pycharm. To better understand the evaluation criterion, we provide four parameters defined in Table 5, where EE+NE+EN+NN=6000 and EE+NE=K.

Based on Table 5, this paper uses Effectiveness, Precision, Recall, and F1-measure as four evaluation criteria to test the recommendation accuracy of TRCRR. The criteria are defined in Table 6.

The proposed TRCRR is compared with KSMT [32], ST4ML [24], and RFS [33] in terms of recommendation accuracy and efficiency. Meanwhile, TRCRR-Cos and TRCRR-Jacc represent the method of applying cosine similarity [34] and Jaccard similarity [35] to TRCRR. The six methods

(TRCRR-Cos, TRCRR-Jacc, TRCRR, KSMT, ST4ML, RFS) are all tested based on the same dataset, experimental procedures, experimental environment, and so on.

5.2 Recommendation accuracy with smaller K

When the number of retrieved topics that users may be interested in is smaller ($K \leq 100$, layer height $h = 1$). The experimental results are shown in Figure 4.

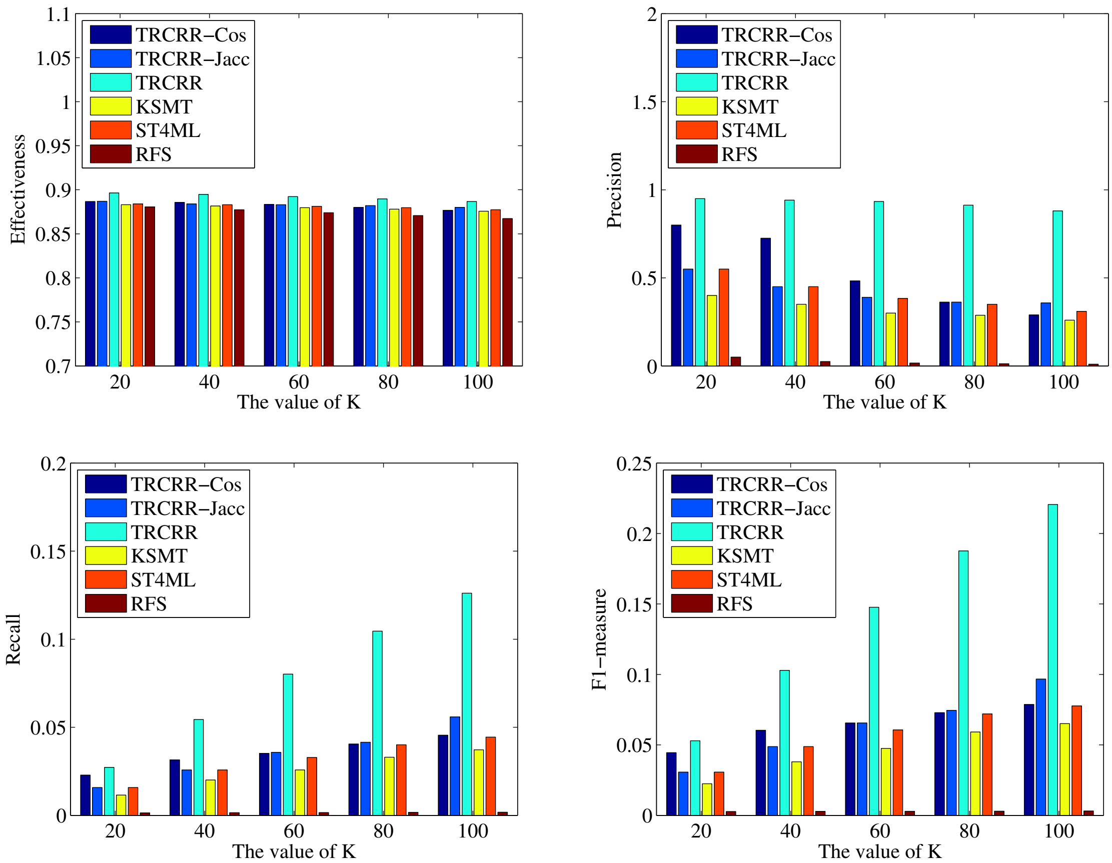

Figure 4: Comparison of the recommendation accuracy of various methods with smaller K.

From Figure 4, it can be seen that our method TRCRR is significantly superior to other methods. As the K value increases, the synthetical recommendation accuracy of the topics recommended by TRCRR is further enhanced. This is because TRCRR utilizes the topic frequency relevancy metric to quantify the relevancy between topics and user preferences. The metric could fully mine the potential correlation between topic-user for enhancing the recommendation accuracy. Using TRCRR-Cos and TRCRR-Jacc to calculate the similarity between topics can excessively refine the gap between topics, making unrelated topics more prominent, and thereby reducing the quality of recommendations. KSMT considers the frequency and contextual relationship between topics by setting co-occurrence windows, which makes it difficult to set appropriate windows and reduces the importance of topic frequency in topic recommendation control. Compared to TRCRR, ST4ML and RFS do not perform well enough, mainly because ST4ML and RFS may not be suitable for regional topic recommendation control.

5.3 Recommendation accuracy with bigger K

When K is bigger ($300 \leq K \leq 1500$), the experimental results are shown in Figure 5. From Figure 5, it can be seen that when the K value increases, the recommendation accuracy of each method decreases to different degrees. The reason for this phenomenon is that as the value of K increases, the proportion of invalid keywords in K gradually increases.

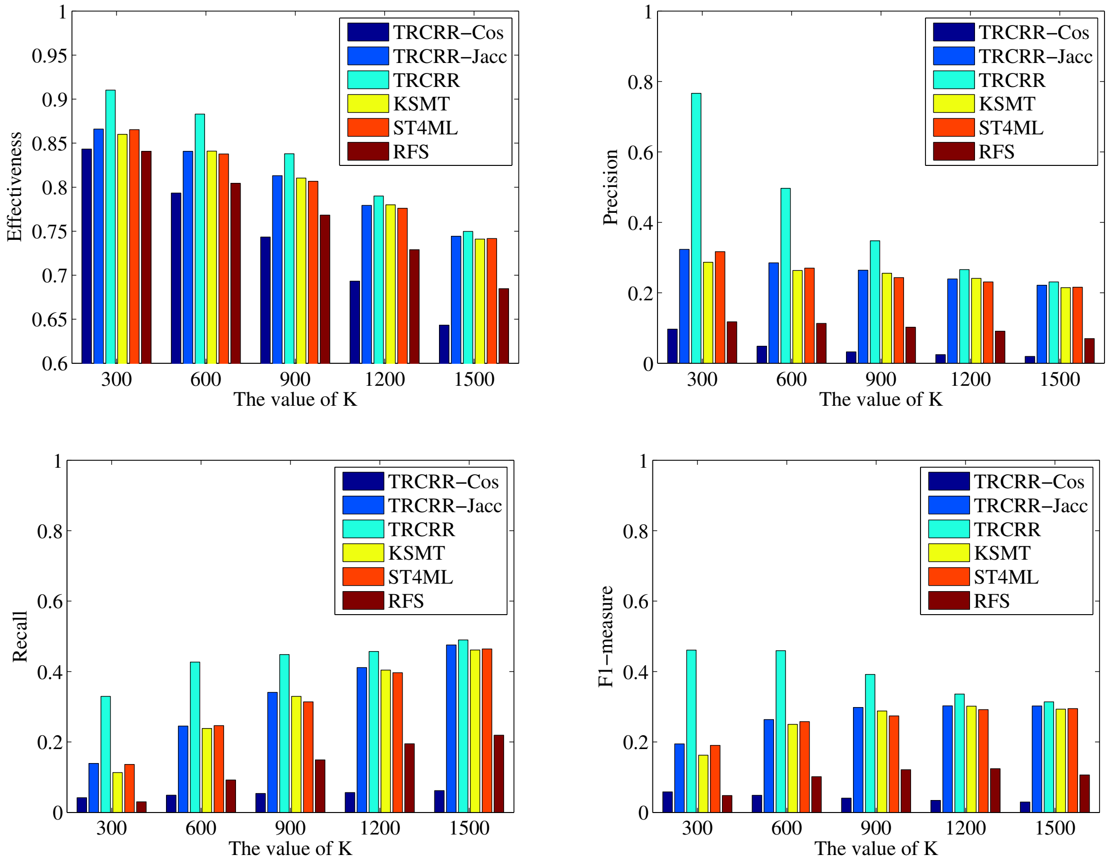

Figure 5: Comparison of the recommendation accuracy of various methods with bigger K.

Among Figure 5, TRCRR still has the best recommendation accuracy. This is because TRCRR tends to place relevant topics at the front of the recommended topics, while topics with reduced relevance are placed at the back of the recommended list. TRCRR-Cos, TRCRR-Jacc, and KSMT distribute relevant topics evenly among the recommended topics, while ST4ML completely intercepts recommended topics based on frequency, which has a certain degree of randomness. RFS distributes relevant topics at the back of the recommended topics.

5.4 Recommendation accuracy with the layer height

As shown in Figure 6, topic recommendation controls are at different layer heights with K=300. TRCRR-Cos, TRCRR-Jacc, and TRCRR extract topics based on their relevancy. These methods require calculating and sorting the relevancy between topics during the recommendation process to extract topics with higher relevancy. The relevancy of topics in KSMT, ST4ML, and RFS is recorded during the text statistics stage, and then the topic list is extracted and integrated based on different sorting criteria and index structures.

By comparison, the first three methods all extract more relevant topics when intercepting a single layer. However, when multiple layers are intercepted and integrated multiple times, the recommendation accuracy decreases, but overall the performance is better, with TRCRR performing the most prominently. The latter three methods, since a large number of invalid topics are filtered out when intercepting topics from layers, have relatively stable recommendation accuracy when multiple layers are intercepted and integrated multiple times. However, compared to TRCRR, their performance is not outstanding enough.

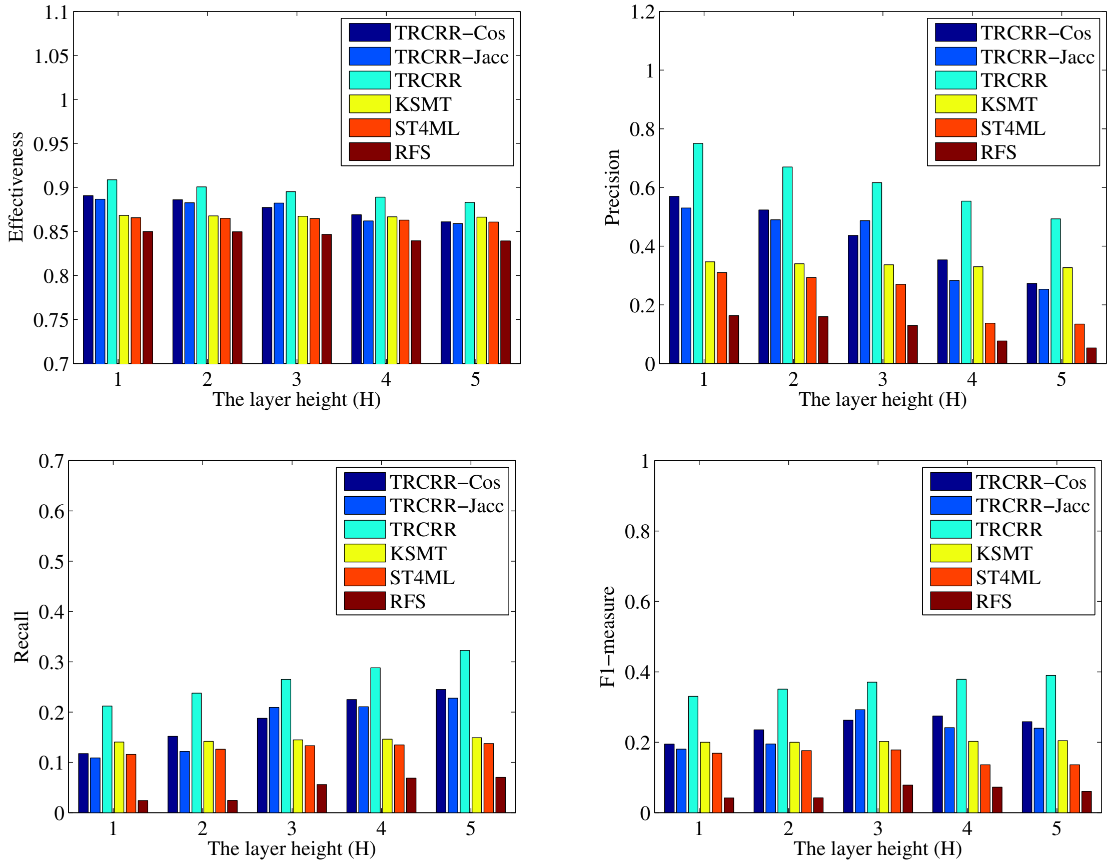

Figure 6: Comparison of the recommendation accuracy of various methods in multi-layer index.

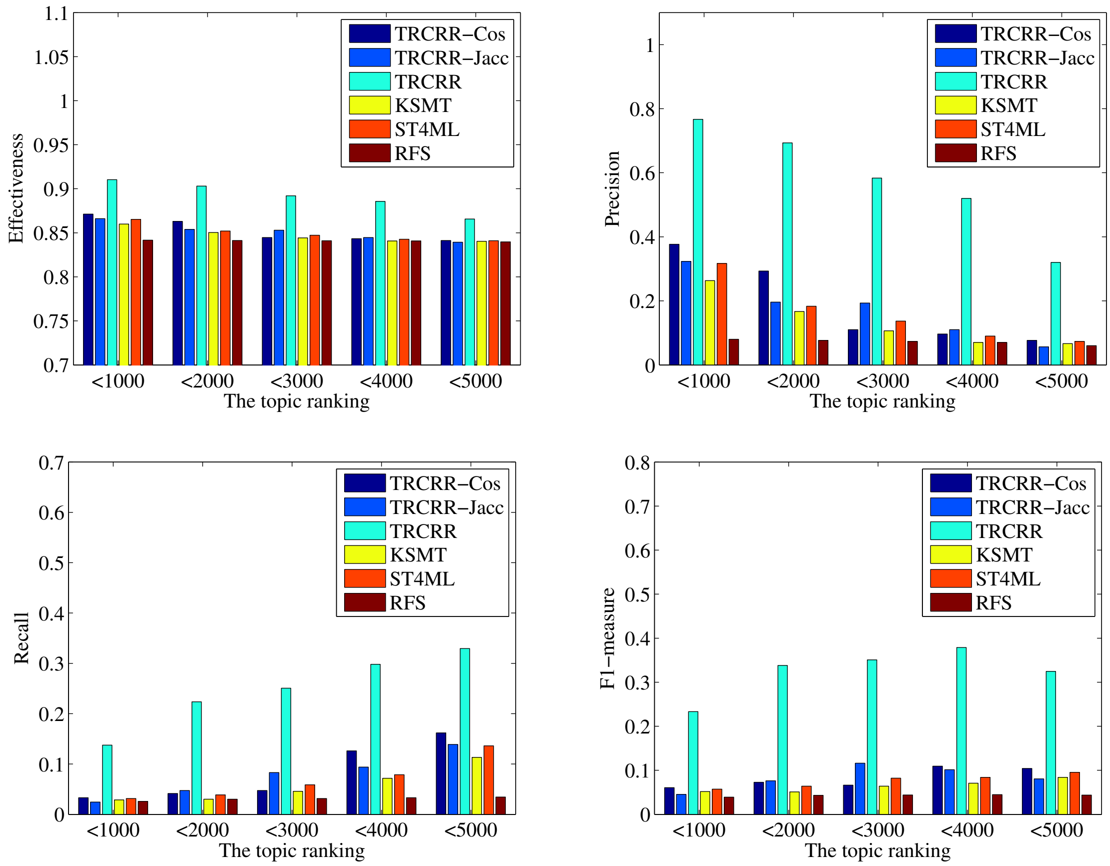

Figure 7: Comparison of recommendation accuracy of various methods under different topic rankings.

5.5 Recommendation accuracy with topic ranking

Figure 7 shows the accuracy comparison (K=300 and H=2) when the ranking of user-interested topics varies in the topic dataset (determined by the number of occurrences). As the ranking of topics that users are interested in decreases in the topic dataset, the recommendation accuracy of topic recommendations decreases accordingly. Among various methods, TRCRR performs the best because it carefully compares the similarity between topics while considering type and frequency information between them. For topics with lower frequency rankings, KSMT and ST4ML methods only consider topics near the frequency ranking as recommended topics. In this case, nearby topics have poor similarity with user-preference topics to a large extent. Moreover, the recommended results displayed by RFS ignore the similarity between higher-ranking topics and lower-ranking topics, resulting in low recommendation accuracy.

5.6 Recommendation efficiency

As shown in Figure 8 (K $\leq$ 100, h=1), compared to KSMT, ST4ML, and RFS, TRCRR-Cos, TRCRR-Jacc, and TRCRR require less time to recommend the same number of topics. This is because when a user provides an interesting topic, KSMT, ST4ML, and RFS first need to find relevant information about the topic from the relevant topic list and record it. Then they determine the topic that needs to be recommended based on the relevant information of the topic. RFS also needs to compare the frequency of various topics in a competitive manner, thus requiring more time. The similarity calculations of TRCRR-Cos and TRCRR-Jacc involve more multiplication and division, while TRCRR uses relatively simple similarity calculations. Therefore, TRCRR requires less time than other methods.

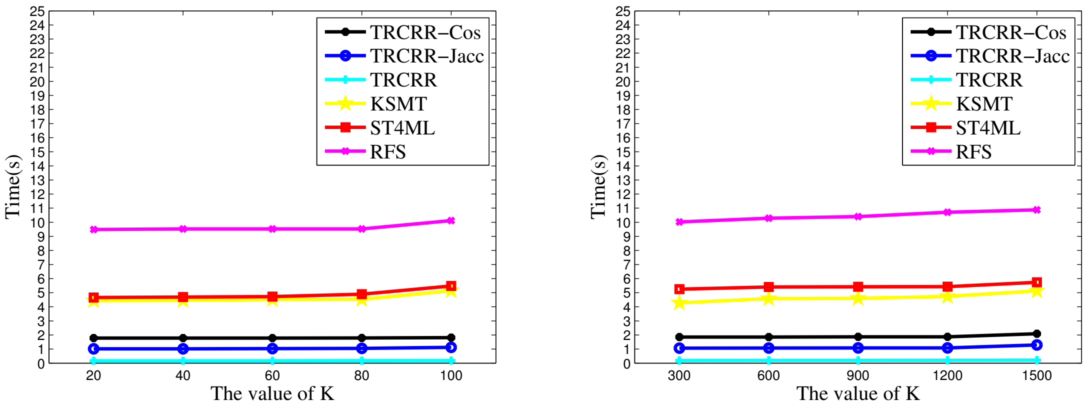

Figure 8: Comparison of time consumption for topic recommendation control under different K values.

5.7 Discussion

Our method TRCRR adopts the personalized topic relevancy list as a measured approach to evaluate the relevancy between topics and user preferences. Thus, our method has better performance than several baseline methods in terms of recommendation accuracy. TRCRR has a lower wall-clock time because of the R-tree topic index, which could reduce the aggregation size of topic lists in different regions to accelerate topic recommendation control.

The relevancy list of TRCRR could be naturally extended to other analytic control methods. So our method has good scalability for more control application scenarios. Meanwhile, the established R-tree index could effectively reduce memory usage. In addition, our R-tree topic index is suitable for data updates. Therefore, it still has good stability when the data gradually increases.

In a word, our method fills the theoretical vacancy of topic recommendation control to some extent. It makes topic recommendation control more suitable for a wide range of big data application scenarios. In addition, our data could be from natural language. That is, our algorithm is related to language processing and could be transferred to language processing [36].

6 Conclusion

In this paper, we proposed TRCRR, a novel method for topic recommendation that leverages user preferences and regional information to improve accuracy and efficiency. The key innovations of TRCRR include a personalized topic relevancy metric that quantifies the similarity between topics and user interests, and an R-tree-based index that organizes topics hierarchically across different regions.

The effectiveness and precision of our method decrease with the increase of K value and layer height. Therefore, we plan to add more factors such as time, social connections, and user feedback into the topic recommendation control. Integrating more factors could potentially further improve the relevance and personalization of the recommended topics. Using tensor algebra to integrate more factors could further improve the effectiveness and precision. Meanwhile, our method does not fully optimize the offline precomputation and indexing steps. Extending TRCRR to handle dynamic updates to the regional topic data incrementally would make it more suitable for real-time streaming scenarios.

Overall, TRCRR represents a promising approach for region-aware topic recommendation that balances accuracy and efficiency. The ideas and techniques introduced in this paper could be valuable not only for direct applications in recommendation systems, but also as building blocks for other problems involving hierarchical spatial data and similarity-based retrieval. We hope that our work will inspire further research on leveraging geographic information to enhance recommendations and other data mining tasks.

Acknowledgment

This work was supported in part by the National Nature Science Foundation of China (Nos. 62362042 and 62262033); in part by the Jiangxi Provincial Social Science Foundation of China (No. 23GL52D); in part by the Key Project of Cognitive Science and Interdisciplinary Research Center of Jiangxi Province (No. RZZD202202); in part by the Annual Project of the "14th Five-Year Plan" for Education Science of Jiangxi Province in 2022 (No. 22YB236); in part by the Foundation of Center for Language Education and Cooperation of the Ministry of Education of China (No. 23YHXZ1001); in part by the Jiangxi Provincial Natural Science Foundation of China (Nos. 20224BAB202012 and 20232BAB202007); and in part by the Scientific and Technological Research Project of Jiangxi Provincial Education Department of China (No. GJJ2201914).

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Luo, S.; Ivison, H.; Han, S. C.; Poon, J. (2024). Local Interpretations for Explainable Natural Language Processing: A Survey, *ACM Computing Surveys*, 56(9), 1-36.

[2] Kim, J.I.; Ju, S.W. (2022). Location Information Analysis of Large Coffee Shops in Big City: A Customer Satisfaction and Behavioral Intention Based Study, *Journal of System and Management Sciences*, 12(1), 63–84.

[3] Rathore, M. A.; Kim, J. (2021). Spatio-temporal Summarized Visualization of SmartX Multi-View Visibility in Cloud-native Edge Boxes, *Computer Science and Information Systems*, 18(1), 169–188.

[4] Christalin Nelson, S.; Tapan Kumar, M.; Prakash G.L. (2022). A Novel Optimized LSTM Networks for Traffic Prediction in VANET, *Journal of System and Management Sciences*, 12(1), 461–479.

[5] Ruchita. (2021). Smart Transportation: Modeling Barriers of Electric Vehicles Charging Infrastructure, *Journal of System and Management Sciences*, 11(4), 190–217.

[6] Čerkauskienė, A.; Meidute-Kavaliauskiene, I. (2023). The Aspects of Supply Chain Risk Management in the Healthcare Industry, *Journal of Logistics, Informatics and Service Science*, 10(1), 1–19.

[7] Buschiazzo M.; Mula J.; Campuzano-Bolarin F. (2020). Simulation Optimization for the Inventory Management of Healthcare Supplies, *International Journal of Simulation Modelling*, 19(2), 255–266.

[8] Peng, H.; Kong, F.; Zhang, Q. (2024). Micro Multiobjective Evolutionary Algorithm With Piecewise Strategy for Embedded-Processor-Based Industrial Optimization, *IEEE Transactions on Cybernetics*, 1–12. DOI: 10.1109/TCYB.2023.3336369

[9] Shamshiri, A.; Ryu, K. R.; Park, J. Y. (2024). Text mining and natural language processing in construction, *Automation in Construction*, 158, 105200.

[10] Anbazhagu, U. V.; Niveditha, V. R.; Bhat, C. R.; Mahesh, T. R.; Swapna, B. (2024). High-Performance Technique for Item Recommendation in Social Networks using Multiview Clustering, *International Journal of Computers Communications & Control*, 19(1), 5818: 1–18.

[11] Stephe, S.; Jayasankar, T.; Kumar, K.V. (2022). Motor Imagery EEG Recognition using Deep Generative Adversarial Network with EMD for BCI Applications, *Tehnički Vjesnik*, 29(1), 92–100.

[12] Alshammari, M.; Alshammari, A. (2023). Friend Recommendation Engine for Facebook Users via Collaborative Filtering, *International Journal of Computers Communications & Control*, 18(2), 4998: 1–10.

[13] Ma, Y.; Jabar, J. (2023). Online Recommendation Method of Malaysian Medical Tourism Products Based on Collaborative Filtering Algorithm, *Journal of Logistics, Informatics and Service Science*, 10(2), 79–90.

[14] Sprenger, B.; De Pasquale, G.; Soloperto, R.; Lygeros, J.; Döfler, F. (2024). Control Strategies for Recommendation Systems in Social Networks, *arxiv preprint arxiv*, https://doi.org/10.48550/arXiv.2403.06152.

[15] Yousfi, H.; Mesmoudi, A.; Hadjali, A.; Matallah, H.; Benkabou, S. E. (2023). SRDF_QDAG: An Efficient End-to-End RDF Data Management When Graph Exploration Meets Spatial Processing, *Computer Science and Information Systems*, 20(4), 1311–1341.

[16] Guo, S.; Liao, X.; Meng, F.; Zhao, Q.; Tang, Y.; Li, H.; Zong, Q. (2024). FSASA: Sequential Recommendation Based on Fusing Session-Aware Models and Self-Attention Networks, *Computer Science and Information Systems*, 21(1), 1–20.

[17] Zhang, X.; Ge, H. E.; Liu, Z.; Liu, Y. Q. (2024). Analysis of Flow Straightener on the Internal Flow Field of Three-Phase Jet Fire Monitors, *International Journal of Simulation Modelling*, 23(1), 101-112.

[18] Kim, K.K.; Park, H.J.; Kim, J.H. (2022). Identification of Herzberg's Motivators and Hygiene Factors for Mobile Shopping Service System Based on Topic Modeling of User Reviews, *Journal of Logistics, Informatics and Service Science*, 9(1), 156-176.

[19] Chen, L.; Li, Y.; Xu, J.; Jensen, C. S. (2018). Towards Why-Not Spatial Keyword Top-k Queries: a Direction-Aware Approach, *IEEE Transactions on Knowledge and Data Engineering*, 30(4), 796-809.

[20] Yin, H.; Wang, W.; Wang, H.; Chen, L.; Zhou, X. (2017). Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation, *IEEE Transactions on Knowledge and Data Engineering*, 29(11), 2537-2551.

[21] Zhou, X.; Mascolo, C.; Zhao, Z. (2019). Topic-enhanced memory networks for personalised point-of-interest recommendation, *In Proceedings of the 25th ACM SIGKDD International conference on knowledge discovery & data mining*, 3018-3028.

[22] Li, R.; Lv, S.; Zhu, H.; Song, X. (2020). Spatial-Temporal Topic Model for Cold-Start Event Recommendation, *IEEE Access*, 8, 214050-214060.

[23] Chen, X.; Liu, Y.; Li, F.; Li, X.; Jia, X. (2021). Remote Sensing Image Recommendation Based on Spatial–Temporal Embedding Topic Model, *Computers & Geosciences*, 157, 104935.

[24] Liu, K.; Tong, P.; Li, M.; Wu, Y.; Huang, J. (2023). ST4ML: Machine Learning Oriented Spatio-Temporal Data Processing at Scale, *Proceedings of the ACM on Management of Data*, 1(1), 1-28.

[25] Zhao, X.; Zhang, Z.; Bi, X.; Sun, Y. (2023). A New Point-of-Interest Group Recommendation Method in Location-Based Social Networks, *Neural Computing and Applications*, 35, 12945–129561.

[26] Chang, W.; Sun, D.; Du, Q. (2023). Intelligent Sensors for POI Recommendation Model Using Deep Learning in Location-Based Social Network Big Data, *Sensors*, 23(2), 850.

[27] Canturk, D.; Karagoz, P.; Kim, S. W.; Toroslu, I. H. (2023). Trust-Aware Location Recommendation in Location-Based Social Networks: a Graph-Based Approach, *Expert Systems with Applications*, 213, 119048.

[28] Gao, Q.; Wang, W.; Huang, L.; Yang, X.; Li, T.; Fujita, H. (2023). Dual-Grained Human Mobility Learning for Location-Aware Trip Recommendation with Spatial–Temporal Graph Knowledge Fusion, *Information Fusion*, 92, 46-63.

[29] Lv, P.; Zhang, Q.; Shi, L.; Guan, Z.; Fan, Y.; Li, J.; Deveci, M. (2024). Exploring on Role of Location in Intelligent News Recommendation from Data Analysis Perspective, *Information Sciences*, 662, 120213.

[30] Wardhani, D.; Astuti, R.; Saputra, D. D. (2024). Optimasi Feature Selection Text Mining: Stemming Dan Stopword Untuk Sentimen Analisis Aplikasi SatuSehat, *Innovative: Journal Of Social Science Research*, 4(1), 7537-7548.

[31] Cai, B.; Shao, Z.; Fang, S.; Huang, X.; Tang, Y.; Zheng, M.; Zhang, H. (2024). The Evolution of Urban Agglomerations in China and How It Deviates from Zipf's Law, *Geo-Spatial Information Science*, 27(1), 38-48.

[32] Chen, Y.; Wang, J.; Li, P.; Guo, P. (2019). Single Document Keyword Extraction via Quantifying Higher-Order Structural Features of Word Co-Occurrence Graph, *Computer Speech & Language*, 57, 98-107.

[33] Farazi, S.; Rafiei, D. (2019). Top-k Frequent Term Queries on Streaming Data, *In 2019 IEEE 35th International Conference on Data Engineering*, 1582-1585.

[34] Hasan, M. R.; Ferdous, J. (2024). Dominance of AI and Machine Learning Techniques in Hybrid Movie Recommendation System Applying Text-to-number Conversion and Cosine Similarity Approaches, *Journal of Computer Science and Technology Studies*, 6(1), 94-102.

[35] Roy, S.; Singh, J.; Ray, S. S. (2024). Weighted Combination of Łukasiewicz Implication and Fuzzy Jaccard Similarity in Hybrid Ensemble Framework (WCLFJHEF) for Gene Selection, *Computers in Biology and Medicine*, 170, 107981.

[36] Nicolae, C. D.; Yadav, R. K.; Tufiş, D. (2023). Evaluation of language models on Romanian XQuAD and RoITD datasets, *International Journal of Computers Communications* & *Control*, 18(1), 5111, 1-15.

Cite this paper as:
Yu J.; Lu Z.; Li X.; Wu B.; Zhang S.; Cui Z. (2024). A Topic Recommendation Control Method Based on User Preferences and Region Communications, *International Journal of Computers Communications* & *Control*, 19(5), 6658, 2024.
https://doi.org/10.15837/ijccc.2024.5.6658