IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 13, 2019, accepted June 15, 2019, date of publication July 1, 2019, date of current version August 31, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2925819*

# Association Rules Mining Among Interests and Applications for Users on Social Networks

**HUAYOU SI[1,2], JIAYONG ZHOU[1,2], ZHIHUI CHEN[1,2], JIAN WAN[1,2], NEAL N. XIONG[3], WEI ZHANG[1,2], AND ATHANASIOS V. VASILAKOS[4]**

[1]School of Computer Science and Technology, Hangzhou Dianzi University, Zhejiang 310018, China
[2]Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Zhejiang 310018, China
[3]Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA
[4]College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China

Corresponding authors: Jian Wan (wanjian@zust.edu.cn) and Neal N. Xiong (xiongnaixue@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grants No. 61472112 and No. 61502129, and the Key R&D Program of Zhejiang Science and Technology Department under Grant 2017C03047.

**ABSTRACT** Interest is an important concept in psychology and pedagogy and is widely studied in many fields. Especially in recent years, the widespread use of many interest-based recommendation systems has greatly promoted research on interest modeling and mining on social networks. However, the existing studies have rarely tried to explore the relationships among interests and their application value, and most similar studies analyze user behavior data. In this paper, we propose and verify two hypotheses about the interests of social network users. We then use association rules to mine users' interests from LinkedIn users' profiles. Finally, based on the interest association rules and user interest distribution on Twitter, we design an approach to mine interests for Twitter users and conduct two experiments to systematically demonstrate the approach's effectiveness. According to our research, we found that there are a large number of association rules between human interests. These rules play a considerable role in our method of interest mining. Our research work not only provides new ideas for interest mining but also reveals the internal relationship between interest and its application value. The research work has certain theoretical and practical value.

**INDEX TERMS** Interests, correlation analysis, association rules, interest mining.

## I. INTRODUCTION

Interests and hobbies refer to individuals' psychological tendencies to desire to know and master something and often participate in such activities or refer to individuals having a cognitive tendency of actively exploring something. In contemporary psychology of interest [1], the term is used as a general concept that may encompass other more specific psychological terms, such as curiosity and to a much lesser degree surprise. [2] In fact, interests have an important influence on personality formation, mental health, education, and career development. They are very important concepts in psychology and pedagogy.

Since the 1980s, scholars have carried out abundant research on interests. In pedagogy, the relationship between interest and teaching is a crucial issue in teaching research and is also an everlasting topic that is always under exploration. For example, Renninger et al. [3] systematically discussed the role of interest in learning and personal development. Hidi and Renninger [4] illustrated the process of interest cultivation. Harackiewicz et al. [5] believe that interest is constructive to academics and that raising interest helps students gain a more proactive learning experience. In psychology, many studies have shown that interest plays a significant role in personality formation and career development, as well as in individual mental health. For example, Sadler et al. [6] studied the changes in students' interests in different periods.

In recent years, with the continuous growth of Internet users and social network applications, the interest-based recommendation systems have been widely used in practice. As a matter of fact, recommending personalized products and information based on user interests and preferences has become a very effective method for product sales and information services. Thus, interest modeling and mining for Internet users and other related research have been

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

gradually carried out. For example, Elmongui et al. [7], Qian et al. [8], Eirinaki et al. [9], and Jiang et al. [10] each proposed a recommendation service method based on user interests. Huang et al. [11], Bhattacharya et al. [12], Zarrinkalam et al. [13], [14], and Li et al. [15] focused on interest modeling for Internet users for different goals and tasks. Moreover, Kapanipathi et al. [16], Xu et al. [17], and Piao and Breslin [18] focused on Interest mining for Internet users based on access logs, microblog/blog accessing, and content and behavior of browsing, respectively. These studies further extend the areas of interest in research, development, and application.

Although research on interests is very extensive, the existing research rarely tries to explore the relationships among interests and their application value based on big data. To address the issue and in combination with the requirements of interest mining for Internet users, we preprocessed the data of LinkedIn and Twitter and at the same time made assumptions and verified their distribution. We then designed a series of methods to mine users' real interests, including obtaining interest relevance and calculating users' sensitivity to interest. Our research work shows there exist many association rules between human interests, which can truly play a very good role in interest mining in our approach. Our contributions in this paper are as follows.

- Based on tens of thousands of profiles with interests from LinkedIn, we analyze the distribution of human interests to mine 210 high frequency interests.
- We analyzed the correlation of interests and then study the association rules among the interests based on our empirical data.
- We analyze the distribution of users' interests on Twitter and demonstrate two hypotheses about the distribution based on empirical data from Twitter and LinkedIn.
- We design an approach to mine interests for Twitter users based on interest association rules and demonstrate the approach's effectiveness.

To facilitate the description of our research, we draw a simple flow chart for mining user interest in social networks, as shown in Figure 1.

The rest of this paper proceeds as follows. Section II discusses the distribution and recognition of interests. Section III studies the association rules among the interests. Section IV analyzes the distribution of users' interests on Twitter. Section V presents our approach for interest mining for Twitter users and then discusses its effectiveness. Section 6 presents related works. The conclusions are drawn in Section 7.

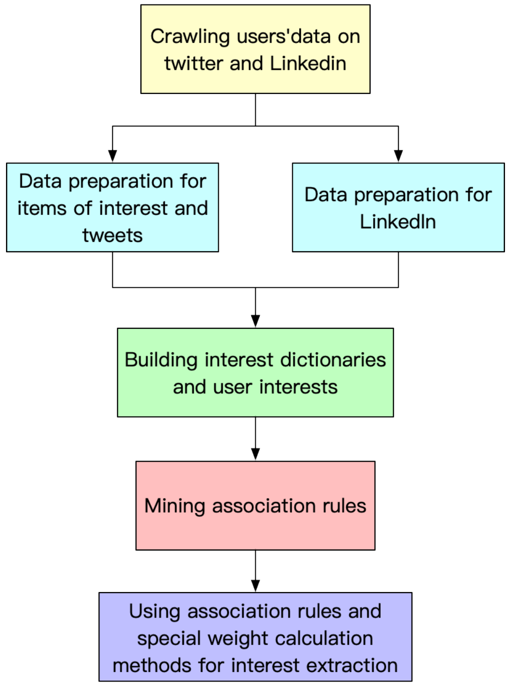



**FIGURE 1.** The flow chart for mining the user's interests.

## II. EMPIRICAL DATA COLLECTION AND INTEREST RECOGNITION

### A. INTEREST DATA COLLECTION

LinkedIn is a very popular business and employment-oriented social networking service. As of September 2016, LinkedIn had more than 467 million accounts. The basic functionality of LinkedIn allows users to create profiles, which typically consist of a curriculum vitae describing their work experience, education and training, interests and hobbies, and a photo of them [19]. The members on LinkedIn usually aim to create a personal professional image, access to business insights, develop professional contacts and find more career opportunities. Compared to other social networking, LinkedIn members can provide more authentic and reliable personal profiles.

LinkedIn members usually list their interests in their profiles. Some interests always appear on the same profiles, which indicates that these interests have an intrinsic connection. For example, the interests "read" and "travel" often appear at the same time. There must be a close relationship between them. Thus, LinkedIn career profiles with interests can be collected to analyze correlation characteristics of interests. In our research, we first design a LinkedIn crawler and then randomly collect 44,623 LinkedIn profiles, of which 10,028 are filled with their interests.

### B. INTEREST RECOGNITION

LinkedIn does not provide a group of interests for its members to choose when they create their profiles. It is very open for members' interests. The members of LinkedIn can freely edit their interests. Therefore, the interests filled in by LinkedIn members are not standardized. In an interest list of a LinkedIn profile, there is no fixed separator between different interests. Some users use the word "*and*", some use a comma ",", some use a semicolon ";", and some directly use a new line to divide different interests. For example, some user's interests are "*Movies and walking*", while some are "*Yoga; hiking; singing; reading; poetry; art; music; Kids*!", which all contain several different interests divided by different separators.

**TABLE 1.** Frequency of parts of interest items in LinkedIn.

| Interest Item | Frequency | Percentage |
|---|---|---|
| travel | 1689 | 3.12% |
| music | 1266 | 2.34% |
| read | 1140 | 2.11% |
| technology | 1073 | 1.98% |
| photography | 811 | 1.50% |
| movie | 772 | 1.43% |
| ski | 742 | 1.37% |
| golf | 674 | 1.25% |
| cycling | 582 | 1.08% |
| running | 551 | 1.02% |
| business | 542 | 1.00% |
| sport | 524 | 0.97% |
| cooking | 491 | 0.91% |
| art | 473 | 0.87% |
| design | 445 | 0.82% |
| family | 427 | 0.79% |

Moreover, LinkedIn users can express the same interest in different words. In natural language, the same interest tends to have a variety of different expressions. Therefore, in this paper we process the interest data collected as follows:

- We first design an algorithm that can intelligently split LinkedIn members' interest list to recognize the interest words as a collection for each user. From the 10,028 profiles with interests, we find 25,913 interest words, which represent respective interests. There is no question that some interest words are synonymous; for example, "ski" and "skiing", "book" and "books", which represent the same interests, are just expressed in different words.
- We then recognize the synonyms and aggregate them into the same interest items. After we proofread artificially, 19430 synonym sets are obtained for all the interest words. There is no question that a synonym set of interest words corresponds to an interest item. To facilitate the description of our work, in this paper, the most frequent interest word in a synonym set is used to name the interest item.
- According to the synonym sets of interest, we replace the interest words in each profile with the names of their own interest items. For each interest item, then we calculate the frequency of its occurrences in 10,028 profiles and the percentage of his occurrences to the total occurrences of all the interest items, which shows the universality of the interest. Parts of the results are shown in Table 1.

From the sorted interest items according to their frequency in descending order, as shown in Figure 2, the cumulative percentage of the top 10 interests is up to 17.02%, the top 50 is up to 37.69%, the top 100 is up to 46.63% . Therefore, we can find that the frequency distribution of the 19430 interests is very uneven, where very few interests have very high frequencies and the frequencies of most of the interests are very low.

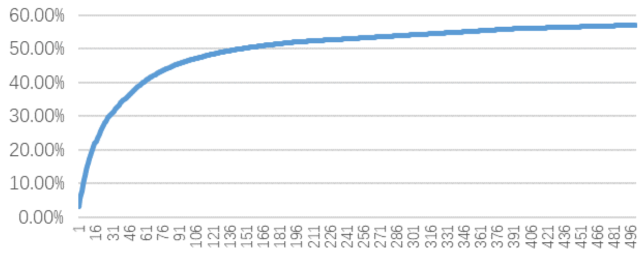



**FIGURE 2.** The cumulative percentage of the Top n interests.

**TABLE 2.** Examples of normalized representation of interests in LinkedIn profiles.

| User ID | Interest Strings Collected | Normalized Representation of Interests |
|---|---|---|
| 168915697 | New technology; Sciences; Languages | technology; language; science |
| 27428582 | Wine, food and good music! | music; food; wine |
| 113724463 | Rugby; Golf; Travel and adventures | travel; golf; rugby; adventure |
| 7735645 | Theater & Improvisation; Tango dancing. | dance; movie |
| 145915690 | Travelling; Football and Fishing | travel; fishing; football |
| 134641445 | Reading; writing; music; eating; cooking; traveling; camping; hiking | read; music; cooking; hike; travel; writing; camping; eating |
| 13715016 | music: piano and guitar; photography (b&w); skiing; badminton | music; photography; ski; guitar; badminton; piano |

- There is no doubt that the higher the frequency of an interest is, the more popular the interest is, and the greater the analytical value is. Therefore, we remove the low-frequency interest items and retain 210 high frequency interest items as subjects of study. In the experimental data, 8,675 out of the 10,028 profiles contain at least one interest in the 210 interest items.

So, for each LinkedIn profile, just keeping the interests in the 210 interest items with standard names, we can get a normalized representation of the interests. Some examples are shown in Table 2.

## III. CORRELATION ANALYSIS FOR INTERESTS

### A. CORRELATION ANALYSIS APPROACH FOR INTERESTS

When something happens in nature, other things will follow. This relationship is called association. The knowledge that reflects dependencies or associations between events is known as relational knowledge. For example, according to shopping basket analysis, some retail rules can be determined, such as "70% of customers who buy a basketball also buy basketball sportswear at the same time" and "40% of all customers buy a basketball and basketball sportswear at the same time". These rules are called association rules. Correlation analysis is also known as association mining, the purpose of which is to find the association rules between data items in a given data set and to describe the degree of closeness between data items. The data set for association rules mining is usually recorded as $D$.

- $D = \{T_1, T_2, \ldots, T_k, \ldots, T_n\}$, where $T_k$ ($k = 1, 2, \ldots, n$) is called a record.

Each record $T_k$ consists of a list of items.

- $T_k = \{i_1, i_2, \ldots, i_m\}$

In this paper, the data record set $D$ refers to the 8,675 profiles with the 210 high frequency interest items. $T_k$ is one of the profiles. The item list of $T_k$ is the collection of interest items in a given profile. In this way, we can establish a correlation analysis approach for interest items.

In correlation analysis, the measurement methods of importance and value for association rules are *confidence*, *support*, *expectation* and *lift*.

- *Confidence*: the measurement of the accuracy and intensity of association rules. The Confidence of the rule $X \to Y$ in data record set $D$ represents the frequency of appearance of $Y$ in all the records where $X$ appears, also representing the inevitability of rule $X \to Y$, denoted as:

$$\begin{aligned} confidence\,(X \to Y) &= P\,(Y\,|\,X) \\ &= |T : X \cap Y \subseteq T, T \in D| \\ &\quad / \,|T : X \subseteq T, T \in D| \times 100\% \end{aligned}$$

- *Support*: the measurement of the importance of association rules, which reflects the universality of association rules and indicates the representation of association rules in all record sets. The Support of rule $X \to Y$ in data record set $D$ represents the frequency of appearance of both $X$ and $Y$ simultaneously in all records, denoted as:

$$\begin{aligned} support(X \to Y) &= P(X \cap Y) \\ &= |\{T : X \cup Y \subseteq T, T \in D\}|/|D| \times 100\% \end{aligned}$$

  where $|D|$ refers to the number of all records in data record set D.
- *Expectation*: for a rule $X \to Y$, it refers to the frequency of the occurrences of $Y$ in all data record sets. In rule $X \to Y$, it describes the frequency of $Y$ in all records sets without any influential factors, denoted as:

$$\begin{aligned} expectation\,(X \to Y) &= P\,(Y) \\ &= |\{T : Y \subseteq T, T \in D|/|D| \times 100\% \end{aligned}$$

- *Lift*: for a rule $X \to Y$, it describes how the occurrence of $X$ affects the appearance of $Y$, which is the ratio of confidence to expectation of the rule, denoted as:

$$\begin{aligned} lift\,(X \to Y) &= P\,(Y\,|\,X)\,/P\,(Y) \\ &= (|T : X \cap Y \subseteq T, T \in D| \times |D|) \\ / \,(|T : X \subseteq T, T \in D| &* |T : Y \subseteq T, T \in D|) \times 100\% \end{aligned}$$

Thus, based on the analysis of interest items, we can mine the association rules among interest items and quantify their characteristics, such as confidence, support, expectation and lift.

**TABLE 3. The numbers of association rules dug out based on different minimum thresholds.**

| Minimum Confidence Threshold | Minimum support threshold | | | | | |
|---|---|---|---|---|---|---|
| | 0.2% | 0.6% | 1% | 1.4% | 1.8% | 2.0% |
| **10%** | 1751 | 309 | 127 | 67 | 38 | 30 |
| **20%** | 857 | 133 | 66 | 37 | 19 | 15 |
| **30%** | 421 | 58 | 21 | 13 | 8 | 5 |
| **40%** | 180 | 17 | 4 | 2 | 1 | 1 |
| **50%** | 86 | 4 | 1 | 1 | 0 | 0 |
| **60%** | 28 | 1 | 0 | 0 | 0 | 0 |
| **70%** | 12 | 1 | 0 | 0 | 0 | 0 |
| **80%** | 3 | 0 | 0 | 0 | 0 | 0 |
| **90%** | 1 | 0 | 0 | 0 | 0 | 0 |

### *B. CORRELATION ANALYSIS OF INTERESTS*

In the mining process of association rules, it is necessary to set the minimum confidence threshold and the minimum support threshold. An association rule that satisfies the thresholds is a strong and meaningful association rule. Apriori [20] is one of the most famous algorithms for mining strong association rules. Based on the empirical data collected in this paper, we apply the Apriori algorithm to mine strong association rules. According to different minimum thresholds, the numbers of strong association rules we dug out are shown in Table 3.

As seen from Table 3, a certain number of interest association rules can be dug out according to different minimum confidence thresholds and minimum support thresholds. Therefore, for the specific requirements in the expected application, a set of strong association rules can be obtained by setting different minimum thresholds.

In addition, some association rules that can be dug out are shown in Table 4. It can be seen from Table 4 that there are some very strong correlations among human interests. An example is the association rule "*culture→travel*", for which the confidence degree is as high as 48.33%, the support degree is as high as 1.16%, and the lift degree is up to 232.77%. This shows that in the human interests, "culture" and "travel" are highly relevant. Another example is the association rule "*read; photography→travel*", for which the confidence degree is 53.24% and the lift degree is 256.43%. Thus, through the empirical correlation analysis, we find that there is a great deal of association relationships among human interests and some association rules have high confidence, lift and support. This shows that there are some intrinsic inherent links among human interests. Therefore, they can be applied to interest mining for users on social networks.

## IV. CHARACTERISTICS OF USER INTERESTS ON TWITTER

### *A. OUR HYPOTHESES*

Twitter is an online social networking service. Users can create accounts on Twitter to post and read short 140-character messages called "tweets". A user's tweets can be spread to that person's followers. At present, Twitter is a very popular user information publishing platform and has more than

**TABLE 4. Examples of interest association rules dug out.**

| No. | Antecedent | Consequent | Confidence | Lift | Support | Expectation |
|---|---|---|---|---|---|---|
| **1.** | friends | family | 59.63% | 983.49% | 1.50% | 6.06% |
| **2.** | culture | travel | 48.33% | 232.77% | 1.16% | 20.76% |
| **3.** | food | travel | 46.67% | 224.78% | 1.13% | 20.76% |
| **4.** | marketing | media | 32.55% | 592.02% | 1.60% | 5.50% |
| **5.** | read; music | movie | 31.05% | 343.53% | 0.99% | 9.04% |
| **6.** | read; photography | travel | 53.24% | 256.43% | 0.85% | 20.76% |
| **7.** | read; cooking | travel | 48.18% | 232.05% | 0.76% | 20.76% |
| **8.** | read; movie | music | 39.09% | 252.88% | 0.99% | 15.46% |
| **9.** | sport; music | travel | 38.01% | 183.09% | 0.75% | 20.76% |
| **10.** | read; movie | travel | 35.00% | 168.59% | 0.89% | 20.76% |

**TABLE 5. Example of interest mining for user *Melgallant* on twitter.**

| Screen_name | *Melgallant* |
|---|---|
| Interests from LinkedIn | media; editing; international relations; writing |
| Ordered Interest List From Twitter | *<media,311>*; <art,261>; <read,104>; <leader,103>; <performance,71>; <Canada,58>; <coffee,49>; <video,43>; <culture,43>; <marketing,43>; <ski,37>; <business,37>; <health,30>; <rock,27>; <internet,27>; <food,25>; <building,25>; <movie,24>; <kids,24>; <design,23>; <UK,20>; <communication,17>; <sport,15>; <surf,14>; <eating,14>; <family,14>; <planning,13>; <talent management,13>; <music,13>; <wine,13>; <dog,13>; <dance,12>; <technology,12>; *<writing,11>*; <drink,11>; <hockey,10>; <law,9>; <bridge,8>; *<editing,8>*; <skate,8>; <analytics,7>; <shopping,7>; <mentor,7>; <nature,7>; <recruitment,6>; <science,6>; <rowing,6>; <sales,6>; <gas,6>; <blogging,6>; <research,6>; <friends,6>; <travel,6>; <innovation,5>; <yoga,5>; <speaking,5>; <shooting,5>; <painting,4>; <security,4>; <startup,4>; <acting,4>; <fashion,4>; <running,4>; <bigdata,3>; <android,3>; <motivation,3>; <fitness,3>; <philanthropy,3>; <camping,3>; <china,3>; <marathon,2>; <risk,2>; <golf,2>; <fishing,2>; *<international relations,2>*; <environment,2>; <opera,2>; <driving,2>; <drums,2>; <cooking,2>; <Asia,2>; <singing,2>; <museum,2>; <India,2>; <oil,2>; <bass,2>; etc. |
| Recall Rate | 100.0% |
| Accuracy Rate | 3.91% |
| F1 Rate | 51.95% |

500 million users. There is no doubt that users are likely to post some tweets that they are interested in. Therefore, we can make the following hypothesis:

- *Hypothesis 1*: The words that can express a Twitter user's interests usually appear in his tweets.

In other words, the interests that are mentioned in the tweets of a Twitter user probably are the interests of the Twitter user. We can even make another hypothesis as follows:

- *Hypothesis 2*: The higher the frequency an interest appears in tweets of a Twitter user, the more likely it is to be the user's real interest.

### B. VERIFICATION OF OUR HYPOTHESES

To verify our hypotheses, from Twitter we first collect the tweets of 930 Twitter users who are all members on LinkedIn and have provided their real interests on LinkedIn. Then, given a user on Twitter, we determine the interests and their frequencies mentioned in all of the user's tweets, where the interests are all the 210 interests dug out in the subsection *Interest Recognition*. There is no doubt that these interests dug from all the tweets of a user are not necessarily his real interests, but his real interest is likely to be among them. For example, from all the tweets of user *Melgallant* on Twitter, we dug out 128 interests with their corresponding frequencies. These interests are sorted by descending order according to their frequencies and shown in Table 5. In addition, we collected his real interests on LinkedIn, which are also shown in Table 5.

In this paper, Recall Rate refers to the percentage of real interests dug out, Accuracy Rate refers to the proportion of real interest in the dug out interests, while F1 Rate refers to the average of Recall Rate and Accuracy Rate. Ordered Interest List from Twitter refers to the list of interests that are dug out for a Twitter user and sorted by descending order according to their frequencies.

In Table 5, we can find that the real interests of user *Melgallant* all appear in his tweets on Twitter, so the recall rate of his interests is 100%. However, we found that his interest rate of accuracy was 3.91%. It can be seen that there are a lot of interests in the tweet that are not his real interest. Moreover, we take the 930 Twitter users with known LinkedIn accounts as empirical samples. We can also find similar results. The specific data are recorded in Table 6.

From Table 6, we can see that the vast majority of users have high recall rates. This means that most of the real interests of the vast majority of users appear in their own tweets. Therefore, we can believe that Hypothesis 1 is true, that is to say, the words that can express a Twitter user's interests usually appear in his tweets.

Then, we will further analyze these data and find that for the empirical samples, 17.42% of the users have highest frequency interests that are their real interests, 15.32% of the users have second interests that are their real interests, 14.00% of the users have third interests that are their real interests, and so on. Parts of the data are shown in Table 7. The Numbers of Users column in Table 7 refers to the numbers of empirical users for which at least a corresponding number of interests can be dug out from their own tweets. For example,

**TABLE 6.** **The statistical results of the empirical samples for recall, accuracy, and F1 rate.**

| Intervals | Recall Rate | | Accuracy Rate | | F1 Rate |
|---|---|---|---|---|---|
| | Numbers of Users | Percentage | Numbers of Users | Percentage | Percentage |
| [100%-90%] | 318 | 34.20% | 1 | 0.11% | 0.00% |
| (90%-80%] | 74 | 7.95% | 0 | 0.00% | 0.00% |
| (80%-70%] | 68 | 7.31% | 0 | 0.00% | 0.00% |
| (70%-60%] | 104 | 11.18% | 1 | 0.11% | 0.00% |
| (60%-50%] | 108 | 11.61% | 4 | 0.43% | 0.11% |
| (50%-40%] | 21 | 2.26% | 2 | 0.22% | 0.54% |
| (40%-30%] | 54 | 5.81% | 0 | 0.00% | 0.54% |
| (30%-20%] | 60 | 6.45% | 12 | 1.29% | 6.23% |
| (20%-10%] | 25 | 2.69% | 99 | 10.65% | 29.57% |
| (10%-0%] | 98 | 10.54% | 811 | 87.21% | 63.01% |

**TABLE 7.** **Proportion of hits of the interests with the Nth highest frequencies.**

| Order of Frequency | Number of Users | Number of Hits | Proportion of Hits |
|---|---|---|---|
| 1 | 930 | 162 | 17.42% |
| 2 | 927 | 142 | 15.32% |
| 3 | 921 | 129 | 14.00% |
| 4 | 915 | 120 | 13.11% |
| 5 | 911 | 103 | 11.30% |
| 6 | 905 | 85 | 9.40% |
| 7 | 902 | 83 | 9.20% |
| 8 | 895 | 76 | 8.50% |
| 9 | 884 | 79 | 8.90% |
| 10 | 871 | 60 | 6.90% |
| 11 | 867 | 67 | 7.70% |
| 12 | 857 | 74 | 8.60% |
| 13 | 850 | 54 | 6.40% |
| 14 | 842 | 51 | 6.00% |
| 15 | 836 | 64 | 7.70% |
| 16 | 829 | 44 | 5.30% |
| 17 | 815 | 57 | 7.00% |
| 18 | 802 | 43 | 5.30% |
| 19 | 794 | 44 | 5.50% |
| 20 | 786 | 40 | 5.10% |

in the tenth row in Table 7, the Number of Users 871 refers to that there are 871 users in empirical samples for whom at least 10 interests are dug out from their own tweets, and the

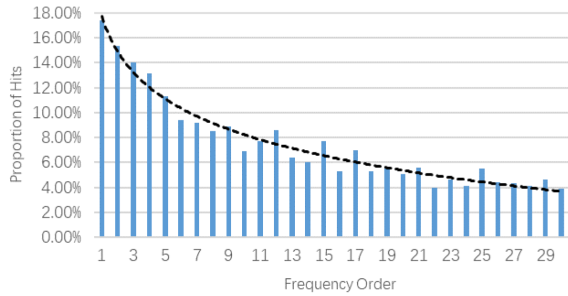



**FIGURE 3.** **Trend of proportion of hits of high frequency of interest.**

Number of hits 60 refers to that there are 60 users in 871 Twitter users whose tenth interest is their own real interest.

From Table 7, we can find that the probability that an interest is a user's real interest usually increases with the increased frequency of that interest in the Twitter user's tweets. Figure 3 depicts the trend of proportions of hits along with the Nth highest frequencies of interests.

## V. INTEREST MINING FOR TWITTER USERS

### *A. OUR APPROACH TO INTEREST MINING*

Although we have confirmed our hypotheses that the words that can express a Twitter user's interests probably appear in his tweets and the higher the frequency of an interest in a user's tweets, the more likely it is to be his real interest, but we cannot distinguish a user's real interest directly from his tweets, since usually numerous interests can be dug out from a Twitter user's tweets. In addition, from Table 7, we can also see that the accuracy rate is very low.

In fact, according to the nature of the interest association rule, we can assume that if a user has an interest, he may also have an interest associated with that interest. Therefore, we apply the interest association rules to reorder the interests of each user that are dug out from Twitter to make their real interests as far as possible appear in the front of the ordered interest list from Twitter. Therefore, we can extract the first few interests as the user's real interests because they are most likely to be the user's real interests.

Without loss of generality, in this paper, we can regard the frequencies of the interests as their weights. For a user's ordered interest list from Twitter, for example in Table 6, we change their weights based on interest association rules and then resort the interests according to their weights by descending order. Therefore, after comprehensive consideration, we designed the following approach to apply interest association rules to interest mining for a user from Twitter, the steps of which are listed below.

1. Given a Twitter user, collect all his tweets from Twitter.
2. According to the 210 high frequency interest items dug out in subsection *Interest Recognition*, mine the interests and their frequencies mentioned in all his tweets.
3. Sort the interests by descending order according to their frequencies as an ordered Interest List from Twitter, denoted as List *oittsList*.

4. Take out all the elements in List *oittsList* as a collection, denoted as Set *ittsSet*.
5. Select a set of interest association rules as Set *ruleSet* dug out in subsection *Correlation Analysis of Interests*.
6. One by one, take out each interest *irt* in List *oittsList*.
7. If there is a rule in Set *ruleSet*, the antecedent of which is interest *irt*, then add its consequents to Set *ittsSet* as interests with the weight $W$.
8. Until each interest in List *oittsList* is processed, sort the interests in Set *ittsSet* according to their weights by descending order to form an interest list, denoted as List *rsltList*.
9. According to the actual needs, take out the first several interests from List *rsltList* as a result of interest mining for the user.

In the process, the weight $W$ is set to $w+k \times r$, where parameter $w$ is an existing weight, if the interests to be added are already in list *ittsSet*; else, parameter $w$ is 0. In this formula, parameter $k$ is the constant used to set the influence of association rules for interest mining in the approach. The greater the value of $k$ is, the greater the influence of association rules for interest mining is. In addition, parameter $r$ is the probability of interest *irt* to be the user's true interest, which refers to the proportion of hits according to the order of interest *irt* in List *oittsList* corresponding to Table 1. This parameter $r$ ensures that if the probability of interest *irt* being the real interest is large, the probabilities of the interests introduced by interest *irt* are large too.

## B. EXPERIMENTAL SETUP FOR EVALUATION

To verify the value of association rules for interest mining, we mine two sets of interest association rules based on different thresholds and then set up two sets of experiments according to the two respective sets of interest association rules for interest mining. Finally, by comparing and analyzing the Proportion of Hits, Recall Rate, Accuracy Rate, and F1 Rate of the results, we determine the value of the association rules for interest mining. The experiment is set up as follows:

- In Experiment 1, we use the association rules dug out through larger thresholds. Therefore, in this experiment, there are fewer association rules, but their association is strong. As shown in Table 6, if the minimum support and confidence thresholds are set to 0.4% and 20% respectively, 286 association rules can be dug out based on our empirical data discussed in Subsection 3.2. In Experiment 1, we take these interest association rules as a set of interest association rules for our approach to interest mining.
- In Experiment 2, we use the association rules dug out through smaller thresholds. Therefore, in this experiment, there are more association rules, but many of them are weak. In the case that the minimum support and confidence thresholds are set to 0.2% and 10%, respectively, we can obtain the 1628 association rules dug out, the lifts of which are all greater than 100%, as a set of interest association rules for interest mining.

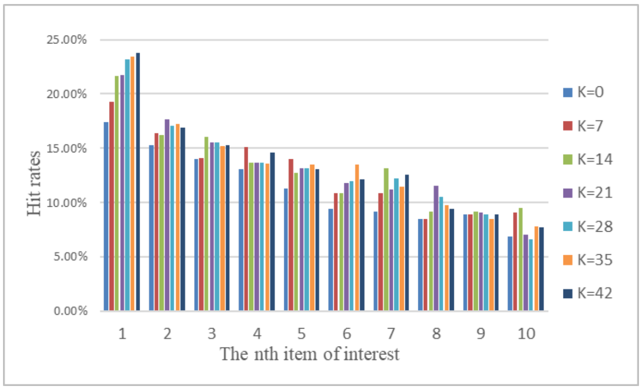



**FIGURE 4.** Comparison of the hit rates in the 7 tests in experiment 1.

In addition, for the two experiments, we apply our approach to process our empirical samples, i.e., the 930 Twitter users discussed in subsection 4.2. Without losing generality, in each experiment, we set our approach's parameter $k$ to 0, 7, 14, 21, 28, 35, and 42 and conduct 7 tests. In fact, when parameter $k$ is set to 0, the association rules do not work, and our approach just returns the original *oittsList* from Twitter as shown in Table 6, where the order of interests is just based on the frequencies of their appearance in users' tweets.

## C. EXPERIMENTAL RESULTS

### 1) EXPERIMENT 1

When the 7 tests are completed in this experiment, for each test, we calculate the proportions of the users whose N-th interests in their own List *rsltList* in our approach are their real interests, which essentially refer to the hit rates of the N-th interests. For example, for each user's first interest in his List *rsltList*, when parameter $k$ is set to 0, 7, 14, 21, 28, 35, and 42, the corresponding proportions are 17.42%, 19.25% 21.61%, 21.72%, 23.23%, 23.44%, and 23.76%, respectively. for ease of understanding, other figures are not explained in detail here. Figure 4 intuitively compares the proportions for the first 10 interests according to the 7 tests.

As seen from Figure 4, once the association rules work, that is, parameter $k$ is not set to 0, the hit rates of the first 10 interest have a certain increase. In some cases, the effect of association rules is obvious. For example, for their first interest, the hit rates are as high as 23.76% when parameter $k$ is set to 42, which is significantly higher than the 17.42% when parameter $k$ is set to 0.Other interest items have similar situations. This means that the application of interest association rules greatly improves the probability that the first several interests in a user's List *rsltList* dug out by our approach are his real interests. This shows that the interest association rules can play a good role in our approach.

Moreover, for each test's results, given a user, we first can get the first 10 interests in his List *rsltList* dug out and calculate the recall rate for him. We then count the proportion of users whose recall rates are greater than a given value. For example, when parameter $k$ is 0, 7, 14, 21, 28, 35, and 42,

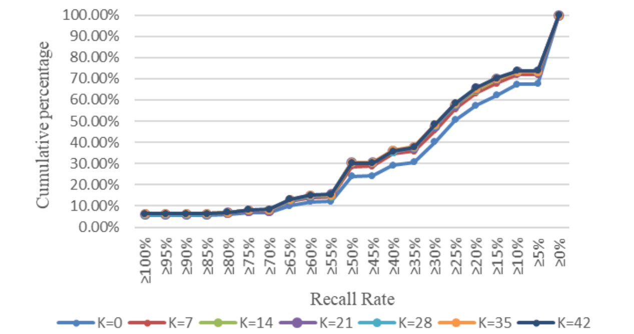



**FIGURE 5.** Comparison of the proportions of users according to different recall rates and parameter k in experiment 1.

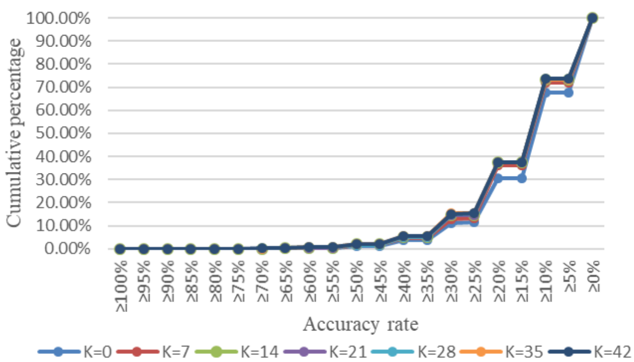



**FIGURE 6.** Comparison of the proportions of users according to different accuracy rates and parameter k in experiment 1.

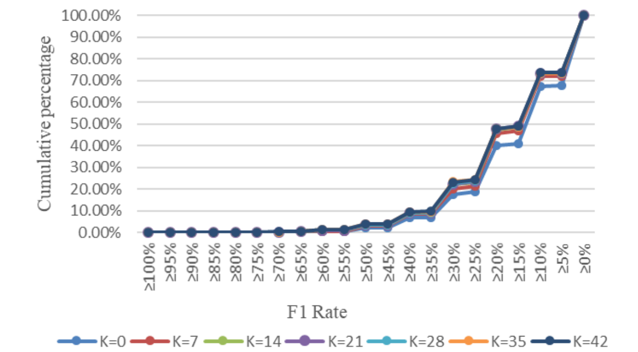



**FIGURE 7.** Comparison of the proportions of users according to different f1 rates and parameter k in experiment 1.

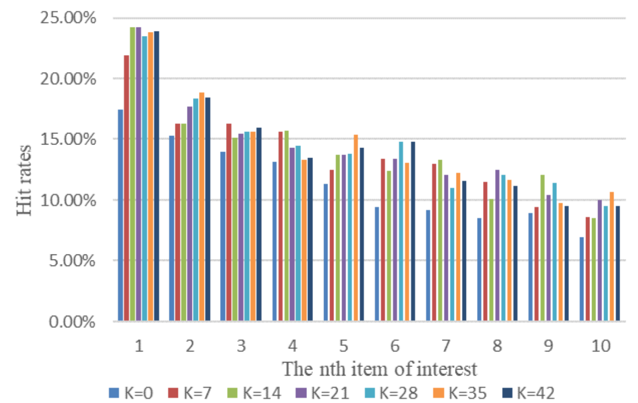



**FIGURE 8.** Comparison of the hit rates in the 7 tests in experiment 2.

the corresponding proportions of users whose recall rates are greater than 70% are 6.77%, 7.96%, 8.17%, 7.85%, 7.96%, 8.28%, and 8.49%, respectively. In another example, the corresponding proportions of users whose recall rates are greater than 30% are 40.22%, 45.59%, 47.42%, 47.42%, 47.96%, 48.49%, and 48.49%. Figure 5 intuitively compares the proportions according to the values of parameter $k$.

From Figure 5, we can find that if the value of parameter $k$ is set to 0, the corresponding curve is the worst. It is obviously inferior to other curves. This means that its recall rate is the lowest under the same conditions. In our tests, if the value of parameter $k$ is set to 0, the corresponding curve is obviously quite good. Here, we can see the association rules obviously improve the recall rate under various weights for application. They have a good value for interest mining. In this experiment, we can see that the greater their weight, the better their effect.

Furthermore, for each test's results and for a user, we first also obtain the first 10 interests in his List *rsltList* that are dug out and calculate the accuracy rate and F1 rate for the user. The proportions of users whose accuracy rates (or F1 rates) are greater than a given value are then counted. For example, when parameter $k$ is set to 0, 7, 14, 21, 28, 35, and 42, the corresponding proportions of users whose accuracy rates are greater than 70% are 0.32%, 0.32%, 0.43%, 0.54%, 0.65%, 0.65%, and 0.65%, respectively. Figure 6 and Figure 7 depict the proportions of users with accuracy rates and F1 rates, respectively, that are greater than a given value.

From Figure 6 and Figure 7, we can also find that the curves corresponding to parameter $k$ with value 0 are inferior to other curves, while the curve corresponding to parameter $k$ with the value of 42 is quite good. This also means that the association rules are valuable for interest mining.

### 2) EXPERIMENT 2

In this experiment, we also conduct the 7 tests just based on the second set of association rules, which has 1628 association rules that are dug out, but many of them are weak. For each test's result, we also calculate the hit rates of the $N^{th}$ interests. For example, for each user's first interest, when parameter $k$ is set to 0, 7, 14, 21, 28, 35, and 42, the corresponding hit rates are 17.42%, 21.94%, 24.19%, 24.19%, 23.44%, 23.76%, and 23.87%, respectively. For each user's second interest, the corresponding hit rate is also over 15%. Figure 8 shows the hit rate of the top 10 interests at different k values.

From Figure 8, we can find that compared to parameter $k$ with value 0, in other cases, the hit rates increase significantly. For example, for their first interest, the hit rates are as high as 24.19% when parameter $k$ is set to 14, which is significantly higher than the 17.42% when parameter $k$ is set to 0. This means that the application of interest association rules greatly

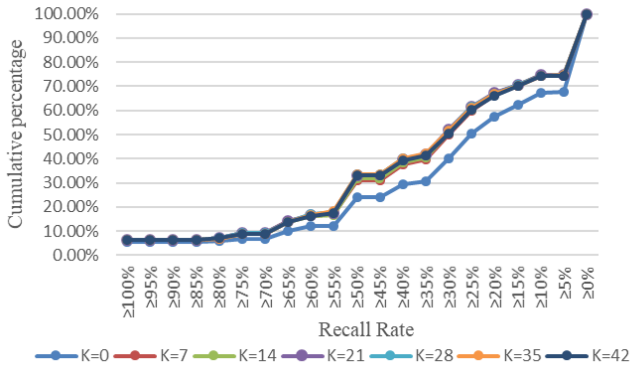



**FIGURE 9.** Comparison of the proportions of users according to different recall rates and parameter k in experiment 2.

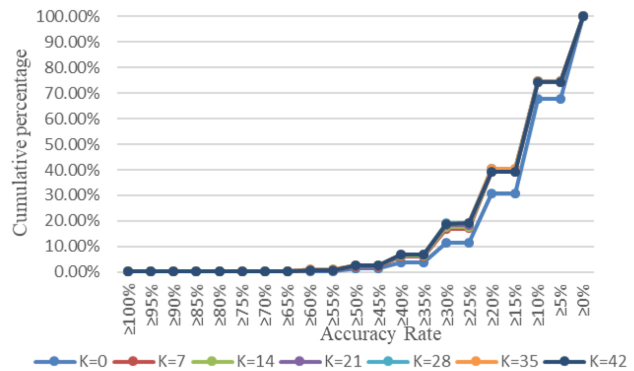



**FIGURE 10.** Comparison of the proportions of users according to different accuracy rates and parameter k in experiment 2.

improves the hit rates of the first several interests in a user's ordered interest list that are dug out. In this experiment, it does not mean that their effect improves as their weight increases.

Furthermore, for each test's results and for a user, we first obtain his first 10 interests that are dug out and calculate the recall rate, accuracy rate, and F1 rate for him. The proportions of users whose recall rates (or accuracy, or rate, F1 rate) are greater than a given value are then counted. To reflect the difference, we set the increment value to 7. For example, when parameter $k$ is set to 0, 7, 14, 21, 28, 35, and 42, the corresponding proportion of users whose recall rates are greater than 70% are 6.77%, 9.25%, 8.92%, 9.03%, 9.14%, 8.92%, and 8.92%, respectively. Figure 9, Figure 10, and Figure 11, respectively, depict the proportions of users according to recall rate, accuracy rate, and F1 rate that are greater than a given value.

From Figure 9, Figure 10, and Figure 11, we also find that the curves corresponding to parameter $k$ with value 0 are very obviously inferior to other curves. This means that the set of association rules are quite valuable for interest mining. When k is set to a different value, the difference between the corresponding curves is not significant. Combined with the first experiment, we believe that not all association rules may be beneficial to mining the real interests of users. In particular, weak association rules may introduce some bias.

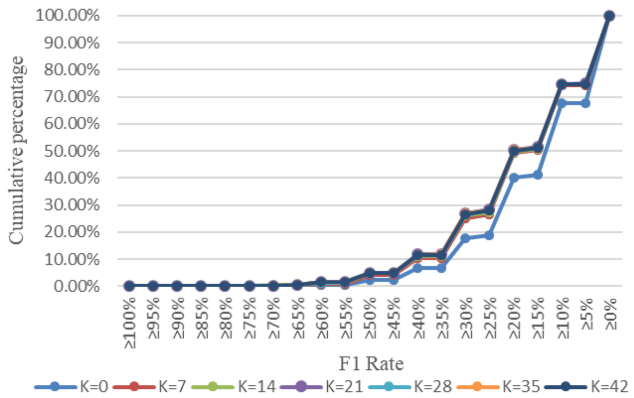



**FIGURE 11.** Comparison of the proportions of users according to different f1 rates and parameter k in experiment 2.

### D. EXPERIMENTAL ANALYSIS

Through the two experiments and their results, we find the interest association rules can truly have a very good effect for interest mining in our approach. As a matter of fact, this conclusion should be reasonable. Since the association rules reflect the relationships between things, in terms of interest, someone has an interest, and to a certain extent, this means that he should have the other interests related to that one.

The results of this experiment can also demonstrate that the interest association rules that we mine based on our empirical big data are reliable because they are valuable for interest mining in our approach.

When parameter k is set to different values, the corresponding values of the recall rate, accuracy rate, and F1 ratio are different. Moreover, in experiment 1, the greater the value of parameter k, the slightly better the effect of association rules is. When the parameter k is set to 28, 35, or 42, the corresponding results are closer. In experiment 2, the value of parameter k has little effect on the experimental results. This means that the set of association rules and their weight in application have subtle effects on interest mining. This is worth exploring further.

In general, the results of experiment 2 are in good agreement with each other. However, they are slightly different, that is, the different sets of rules are slightly different in the application. We can further argue that the minimum confidence thresholds and minimum support thresholds for the association rules mining will influence the expected application. This is worth exploring further too.

## VI. RELATED WORKS

Interests are very important concepts in psychology and pedagogy. Since the 1980s, scholars have carried out considerable amounts of research on interests in different areas of research. Michelson and Macskassy [21] use a knowledge base to eliminate and classify the ambiguities of entities in Tweets. They then develop a "topic profile", which characterizes users' topics of interest, by discerning which categories appear frequently and cover the entities. In pedagogy, Renninger et al. [3] illustrate the role of interest in learning and personal development. They agree that interest is an important force to promote learning. Therefore, it is very meaningful to cultivate and improve students' interest in teaching. Hidi and Renninger [4] systematically study the cultivation of interests. They elaborate on the four-stage interest cultivation process. Harackiewicz et al. [5] proposed a four-stage model of interest development, which can help to establish some measures to effectively enhance interest. In recommendation, Phelan et al. [22] describe a new approach to news recommendations that uses real-time microblogging activity from services such as Twitter as the basis for promoting news stories from users' favorite RSS feeds. Sriram et al. [23] provide a short text classification method. They propose using a small set of domain-specific features extracted from the author's profile and text. The proposed approach effectively classifies the text to a predefined set of generic classes such as News, Events, Opinions, Deals, and Private Messages. In addition, Sadler [6] observed changes in the interests of more than 6,000 students in common occupations at different times.

In recent years, along with the development of the Internet, the interest-based recommendation systems have been widely used in e-commerce and social networking. Thus, interest modeling and mining for Internet users have been gradually carried out. For example, Elmongui et al. [7] proposed a personalized recommendation system for the user's timeline that combines his user characteristics, social behavioral characteristics and tweet content to capture his interests. Qian et al. [8] design a unified personalized recommendation model based on personal interest, interpersonal interest similarity, and interpersonal influence. The factor of personal interest can make the recommended items meet users' individualities, especially for experienced users. For the cold start users, the interpersonal interest similarity and interpersonal influence can enhance the intrinsic link among features in the latent space. Their experimental results show the proposed approach outperforms the main existing approaches. Eirinaki et al. [9] proposed a model user interest community detection model to analyze the text flow from the Weibo website to detect the user's interest community. His user interest model can solve the problem that existing community detection methods ignore the structural and semantic information of posts. In addition, an allocation model is proposed, which is based on improved hypertext-induced topic search, which can reduce the negative impact of non-related users and their interests to improve the accuracy of extracting interest and high-impact users. The experimental results prove that this model can effectively solve the sparsity problem of user interest community detection and solving post data. In addition, Vijayakumar et al. [24], Yin et al. [25], Zhao et al. [26], Xu [27] and other scholars have also put forward their own methods in this research area. Moreover, Vijayaraghavan et al. [28] and Yee et al. [29] have applied for U.S. Patents for their interest-based recommendation systems.

For interest modeling and its application, Zarrinkalam et al. [13] integrate the temporal evolution of semantic information and user interests from the Wikipedia category structure into their predictive models to address the limitations of existing methods of interest space operations. Specifically, in order to capture the temporal behavior of the topic and the user's interests, they consider discrete intervals and construct the user's topic profile in each time interval. Then, the interests observed by the user over several time intervals are summarized by transferring them over the Wikipedia category structure. The experimental results show that they not only enable us to summarize the interests of users but also enable us to transfer users' interests at different time intervals that do not necessarily have the same set of topics. Bhattacharya et al. [12] propose KAURI, a graph-based framework to collectively link all the named entities in all tweets posted by a user via modeling the user's topics of interest. They argue that each user has a potential distribution of thematic interests across the various named entities, and then combines the interest information associated with the user information associated with the tweets into a unified graph based framework. Their experimental results show that KAURI significantly outperforms the baseline methods in terms of accuracy. Zarrinkalam et al. [14] argue that existing methods of identifying user interest rely heavily on explicit contributions (posts) from users, ignoring implicit user interest, that is, potential users who are not explicitly mentioned but may be interested. So he proposed a prediction model based on graph join, which runs on a representation model composed of three types of information: the explicit contribution of users to the topic, the relationships between users, and the relevance of topics. The comparison of the real-world Twitter public demo dataset shows that this model is very effective in building a cold-start user interest file. In addition, in order to solve the problem that the SATM model is too strict and consumes a large-scale corpus, Li et al. [15] propose a generalized topic model (LTM) for short text, provided that the observable short text is generated from the original document. The membership of the original document is unknown. Experimental results show that the model is more competitive than commonly used models. Huang et al. [11] built a user model of heterogeneous networks with undirected and directed edges and applied the model to propose a new approach to overlapping community detection in heterogeneous social networks (OCD-HSN). Compared with the existing state-of-the-art algorithms, this method shows higher accuracy and lower time consumption under the real social network.

In terms of interest mining, Kapanipathi et al. [16] establish a hierarchy-based semantics system that infers user interests expressed as hierarchical interest graphs by leveraging the hierarchical relationships existing in the knowledge base and then uses different levels of conceptual abstraction to personalize or recommend projects. The results show that this method is effective for the users we study. Xu [17] proposed a new unsupervised learning model-latent interest and topic mining model (LITM), which is used to automatically mine latent user interests and project topics from the user-project bipartite network. Experiments show that this work can effectively alleviate the limitations of a latent factor model (LFM), and the experimental results verify the effectiveness of LITM model training and its ability to provide better service recommendation performance based on a user-project binary network. In addition, He et al. [30], Deng et al. [31] and other scholars have also proposed their own methods for interest mining. Based on user preferences, Zhou et al. [32] design a two-stage mining algorithm (GAUP) to mine the most influential nodes in a network on a given topic. Given a set of users' documents labeled with topics, GAUP first computes user preferences with a latent feature model based on SVD or a model based on vector space and then finds Top-K nodes in the second stage. Overall, these approaches for interest mining for Internet users are based on access logs, microblog/blog accessing, and content and behavior of browsing.

In the larger context, in recent years, social network data mining has been extensively studied. However, extracting intelligence from such data has become a quickly widening multidisciplinary area that demands the synergy of scientific tools and expertise. Sapountzi A and Psannis K E [33] illustrate the entire spectrum of social data networking analysis and their associated frameworks and provide a sophisticated classification of state-of-the-art frameworks considering the diversity of practices, methods and techniques. They demonstrate challenges and future directions with a focus on text mining and the promising avenue of computational intelligence. Zhou et al. [34] concentrate on user role identification based on their social connections and influential behaviors in order to facilitate information sharing and propagation in social networking environments. Chen et al. [35] present a study of deceptive information of great benefit to the detection of Twitter spam. Guo et al. [36] propose a novel method for crawling to extract fresh information from online social networks in an efficient and effective manner. Moreover, the interest mining for users has a wide range of application prospects, such as travel recommendation [37], user personality analysis [38], organizational behavior analysis [39], and so on [40]. However, just for interest mining, existing research work being consulted rarely involves the inner relationship among interests and its application.

## VII. CONCLUSIONS

Based on a large amount of empirical data from social networks, in this paper we have performed the following four research tasks.

- Collecting tens of thousands of profiles with personal interests from LinkedIn as our empirical data, we analyze the distribution of human interests and then mine 210 high frequency interests as the objects of study.
- We analyzed the correlation of interests and study the association rules among the 210 interests based on our empirical data.
- Based on hundreds of Twitter users with known interests, we analyze the distribution characteristics of users' interests on Twitter.
- Based on interest association rules and users' interest distribution on Twitter, we design an approach to interest mining for Twitter users and demonstrate the approach's effectiveness.

According to our studies in this paper, we figured out that there exists a large number of correlations between human interests, and some association rules have very high degrees of confidence, lift and support. These findings show that there are some inherent fixed relationships among human interests. In addition, we find that when the interest association rules are applied to interest mining, they can truly play a very good role in interest mining in our approach.

Our research work not only provides a new idea for interest mining but also reveals the intrinsic relationships of association and dependency among interests and their application value. In fact, the research work has considerable theoretical and practical value.

In this research work, we also found some topics that are worth exploring further. Soon, we will carry out the following research work.

a) Study the optimal solution in which association rules apply to interest mining, such as the choice of rule sets and the setting of their weights.
b) Empirically analyze the clustering relationships among interests based on big data and study their application value in interest mining.

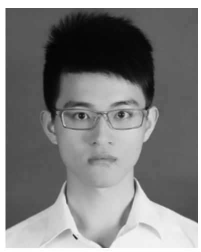
In addition, we will apply the related theory and methods in other areas of research, such as the theories [38], [39], to study relationships among users in social networking platform Twitter. Moreover, we will improve the capabilities of data processing in our approach to promote practicality for large-scale data sets.

## REFERENCES

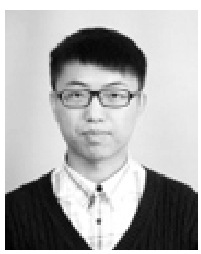
[1] P. J. Silvia, *Exploring the Psychology of Interest*. New York, NY, USA: Oxford Univ. Press, 2006.
[2] *Interest (Emotion)*, Wikipedia, Mar. 2019.
[3] K. A. Renninger, S. Hidi, A. Krapp, and A. Renninger, *The Role of interest in Learning and Development*. New York, NY, USA: Psychology Press, 2014.
[4] S. Hidi and K. A. Renninger, "The four-phase model of interest development," *Edu. Psychol.*, vol. 41, no. 2, pp. 111–127, 2006.
[5] J. M. Harackiewicz, J. L. Smith, and S. J. Priniski, "Interest matters: The importance of promoting interest in education," *Policy Insights Behav. Brain Sci.*, vol. 3, no. 2, pp. 220–227, Oct. 2016.
[6] P. M. Sadler, G. Sonnert, Z. Hazari, and R. Tai, "Stability and volatility of STEM career interest in high school: A gender study," *Sci. Educ.*, vol. 96, no. 3, pp. 411–427, 2012.
[7] H. G. Elmongui, R. Mansour, H. Morsy, S. Khater, A. El-Sharkasy, and R. Ibrahim, "TRUPI: Twitter recommendation based on users' personal interests," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, 2015, pp. 272–284.

[8] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1763–1777, Jul. 2014.
[9] M. Eirinaki, J. Gao, I. Varlamis, and K. Tserpes, "Recommender systems for large-scale social networks: A review of challenges and solutions," *Future Gener. Comput. Syst.*, vol. 78, pp. 413–418, Jan. 2018.
[10] L. Jiang, L. Shi, L. Liu, J. Yao, and M. A. Yousuf, "User interest community detection on social media using collaborative filtering," in *Wireless Networks*. New York, NY, USA: Springer, 2019.
[11] M. Huang, G. Zou, B. Zhang, Y. Liu, Y. Gu, and K. Jiang, "Overlapping community detection in heterogeneous social networks via the user model," *Inf. Sci.*, vol. 432, pp. 164–184, Mar. 2018.
[12] P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi, "Inferring user interests in the twitter social network," in *Proc. 8th ACM Conf. Recommender Syst.*, 2014, pp. 357–360.
[13] F. Zarrinkalam, H. Fani, E. Bagheri, and M. Kahani, "Predicting users' future interests on Twitter," in *Proc. Eur. Conf. Inf. Retr.*, 2017, pp. 464–476.
[14] F. Zarrinkalam, M. Kahani, and E. Bagheri, "Mining user interests over active topics on social networks," *Inf. Process. Manage.*, vol. 54, no. 2, pp. 339–357, Mar. 2018.
[15] X. Li, C. Li, J. Chi, and J. Ouyang, "Short text topic modeling by exploring original documents," *Knowl. Inf. Syst.*, vol. 56, no. 2, pp. 443–462, 2018.
[16] P. Kapanipathi, P. Jain, C. Venkataramani, and A. Sheth, "User interests identification on twitter using a hierarchical knowledge base," in *The Semantic Web: Trends and Challenges*. Crete, Greece: ESWC, 2014, pp. 99–113.
[17] J. Xu, S. Wang, S. Su, S. A. P. Kumar, and C. Wu, "Latent interest and topic mining on user-item bipartite networks," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jun./Jul. 2016, pp. 778–781.
[18] G. Piao and J. G. Breslin, "Inferring user interests for passive users on twitter by leveraging followee biographies," in *Proc. Eur. Conf. Inf. Retr.*, 2017, pp. 122–133.
[19] *LinkedIn—Wikipedia*. Accessed: Mar. 30, 2019. [Online]. Available: https://en.wikipedia.org/wiki/LinkedIn
[20] P. Asthana, A. Singh, and D. Singh, "A survey on association rule mining using Apriori based algorithm and hash based methods," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 7, pp. 599–603, 2013.
[21] M. Michelson and S. A. Macskassy, "Discovering users' topics of interest on twitter: A first look," in *Proc. 4th Workshop Anal. Noisy Unstructured Text Data*, 2010, pp. 73–80.
[22] O. Phelan, K. McCarthy, and B. Smyth, "Using twitter to recommend real-time topical news," in *Proc. 3rd ACM Conf. Recommender Syst.*, 2009, pp. 385–388.
[23] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2010, pp. 841–842.
[24] V. Vijayakumar, S. Vairavasundaram, R. Logesh, and A. Sivapathi, "Effective knowledge based recommender system for tailored multiple point of interest recommendation," *Int. J. Web Portals*, vol. 11, no. 1, pp. 1–18, 2019.
[25] H. Yin, B. Cui, Z. Huang, W. Wang, X. Wu, and X. Zhou, "Joint modeling of users' interests and mobility patterns for point-of-interest recommendation," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 819–822.
[26] S. Zhao, I. King, and M. R. Lyu, "Aggregated temporal tensor factorization model for point-of-interest recommendation," *Neural Process Lett.*, vol. 47, no. 3, pp. 975–992, Jun. 2018.
[27] K. Xu, X. Zheng, Y. Cai, H. Min, Z. Gao, B. Zhu, H. Xie, and T.-L. Wong, "Improving user recommendation by extracting social topics and interest topics of users in Uni-directional social networks," *Knowl.-Based Syst.*, vol. 140, pp. 120–133, Jan. 2018.
[28] R. Vijayaraghavan, S. R. Kulkarni, and K. M. Adusumilli, "Intent prediction based recommendation system using data combined from multiple channels," U.S. Patent 20 170 213 274 A1, Jul. 27, 2017.
[29] Y. H. Yee, J. V. McFadden, J. Kraemer, and D. Sampath, "Methods, systems, and media for recommending content items based on topics," U.S. Patent 20 170 103 343 A1, Apr. 13, 2017.
[30] L. He, Y. Jia, W. Han, and Z. Ding, "Mining user interest in microblogs with a user-topic model," *China Commun.*, vol. 11, no. 8, pp. 131–144, Aug. 2014.
[31] L. Deng, Y. Jia, B. Zhou, J. Huang, and Y. Han, "User interest mining via tags and bidirectional interactions on Sina Weibo," *World Wide Web*, vol. 21, no. 2, pp. 515–536, Mar. 2018.
[32] J. Zhou, Y. Zhang, and J. Cheng, "Preference-based mining of top-*K* influential nodes in social networks," *Future Gener. Comput. Syst.*, vol. 31, pp. 40–47, Feb. 2014.
[33] A. Sapountzi and K. E. Psannis, "Social networking data analysis tools & challenges," *Future Gener. Comput. Syst.*, vol. 86, pp. 893–913, Sep. 2016.
[34] X. Zhou, B. Wu, and Q. Jin, "User role identification based on social behavior and networking analysis for information dissemination," *Future Gener. Comput. Syst.*, vol. 96, pp. 639–648, Jul. 2017.
[35] C. Chen, S. Wen, J. Zhang, Y. Xiang, J. Oliver, A. Alelaiwi, and M. M. Hassan, "Investigating the deceptive information in Twitter spam," *Future Gener. Comput. Syst.*, vol. 72, pp. 319–326, Jul. 2017.
[36] R. Guo, H. Wang, M. Chen, J. Li, and H. Gao, "Parallelizing the extraction of fresh information from online social networks," *Future Gener. Comput. Syst.*, vol. 59, pp. 33–46, Jun. 2016.
[37] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 151–158, Feb. 2016.
[38] S. Laumer, C. Maier, A. Eckhardt, and T. Weitzel, "User personality and resistance to mandatory information systems in organizations: A theoretical model and empirical test of dispositional resistance to change," *J. Inf. Technol.*, vol. 31, no. 1, pp. 67–82, 2016.
[39] M. J. Gelfand, Z. Aycan, M. Erez, and K. Leung, "Cross-cultural industrial organizational psychology and organizational behavior: A hundred-year journey," *J. Appl. Psychol.*, vol. 102, no. 3, pp. 514–529, 2017.
[40] W. Gao, J. L. G. Guirao, B. Basavanagoud, and J. Wu, "Partial multi-dividing ontology learning algorithm," *Inf. Sci.*, vol. 467, pp. 35–58, Oct. 2018.

**HUAYOU SI** received the M.S. and Ph.D. degrees in computer science from Peking University, in 2004 and 2012, respectively. He is currently a Lecturer with the School of Computer Science and Technology, Hangzhou Dianzi University. During the past several years, his research interests include P2P networks, service-oriented computing, and semantic web. In the related research field, he has published more than 20 academic papers. In addition, he has served in the technical program committee for several international conferences.



**JIAYONG ZHOU** was born in Hangzhou, China. He received the bachelor's degree in computer science and technology from the Zhejiang Agriculture and Forestry University and Yangyang College, in 2017. He is currently pursuing the master's degree in computer technology with Hangzhou Dianzi University. His current interests are data mining and the application of big data in the social field.



**ZHIHUI CHEN** received the B.Eng. degree from the School of Computer Science and Technology, Hangzhou Dianzi University, China, in 2015, where he is currently pursuing the master's degree with the School of Computer Science and Technology. His research interests include data mining and web service. In addition, he is a Student Member of the China Computer Federation (CCF).

**JIAN WAN** received the Ph.D. degree in computer technology from Zhejiang University. He is currently a Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, China. His research interests include virtual computing, grid computing, service computing, and embedded systems. He is a member of the Association for Computing Machinery (ACM) and the China Computer Federation (CCF).

**NEAL N. XIONG** received the Ph.D. degree from Wuhan University (in sensor system engineering), and the Ph.D. degree from the Japan Advanced Institute of Science and Technology (on dependable sensor networks).

He is currently an Associate Professor (3rd year) with the Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA. Before he attended Northeastern State University, he worked at Georgia State University, Wentworth Technology Institution, and the Colorado Technical University (Full Professor for approximately five years) for approximately 10 years. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory. He has published over 280 international journal papers and over 120 international conference papers. Some of his works were published in the IEEE JSAC, IEEE or ACM transactions, ACM Sigcomm workshop, the IEEE INFOCOM, ICDCS, and IPDPS. He has been the General Chair, the Program Chair, the Publicity Chair, a PC member, and an OC member of over 100 international conferences and as a Reviewer for approximately 100 international journals, including the IEEE JSAC, the IEEE SMC (Park: A/B/C), the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON MOBILE COMPUTING.

**WEI ZHANG** received the B.E. degree from the School of Information Science and Engineering, Wuhan University of Science and Technology, China, in 2000, and the M.Ec. and Ph.D. degrees from the Computer School, Wuhan University, China, in 2004 and 2008, respectively. He is currently an Associate Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, China. His research interests include wireless sensor networks and intelligent computing. He is a member of the Association for Computing Machinery (ACM) and the China Computer Federation (CCF).

**ATHANASIOS V. VASILAKOS** is currently a Professor with the Department of Computer and Telecommunications Engineering, University of Western Macedonia, Greece, and also a Visiting Professor at the Graduate Programme of the Department of Electrical and Computer Engineering, National Technical University of Athens (NTUA). He has coauthored (with W. Pedrycz) many books, namely the *Computational Intelligence in Telecommunications Networks* (CRC Press, USA, 2001), the *Ambient Intelligence*, the *Wireless Networking*, and the *Ubiquitous Computing* (Artech House, USA, 2006); has coauthored (with M. Parashar, S. Karnouskos, W. Pedrycz) the *Autonomic Communications* (Springer), and the *Arts and Technologies* (MIT Press); and has coauthored (with M. Anastasopoulos) the *Game Theory in Communication Systems* (IGI Inc., USA). He has published more than 150 articles in top international journals and conferences. He is the Chairman of the Telecommunications Task Force of the Intelligent Systems Applications Technical Committee (ISATC) of the IEEE Computational Intelligence Society (CIS). He is the Senior Deputy Secretary-General and a Fellow Member of the ISIBM (International Society of Intelligent Biological Medicine. He is a member of the ACM. He is the Editor-in-Chief of the Inderscience Publishers journals: the *International Journal of Adaptive and Autonomous Communications Systems* and the *International Journal of Arts and Technology* (IJART). He has been on the editorial board of more than 20 international journals, including the *IEEE Communications Magazine* (1999–2002 \& 2008–current), the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS (SMC, Part B, 2007–current), the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (invited), and the *ACM Transactions on Autonomous and Adaptive Systems* (invited).