Hindawi
Journal of Healthcare Engineering
Volume 2021, Article ID 3361755, 9 pages
https://doi.org/10.1155/2021/3361755

Research Article

Management and Plan of Undergraduates' Mental Health Based on Keyword Extraction

Weifeng Zhang

Xinxiang University, Xinxiang, Henan 453003, China

Correspondence should be addressed to Weifeng Zhang; zhangweifeng02@xxu.edu.cn

Received 23 September 2021; Revised 6 October 2021; Accepted 11 October 2021; Published 28 October 2021

Academic Editor: yan chai hum

Mental health issues are alarmingly on the rise among undergraduates, which have gradually become the focus of social attention. With the emergence of some abnormal events such as more and more undergraduates' suspension, and even suicide due to mental health issues, the social attention to undergraduates' mental health has reached a climax. According to the questionnaire of undergraduates' mental health issues, this paper uses keyword extraction to analyze the management and plan of undergraduates' mental health. Based on the classical TextRank algorithm, this paper proposes an improved TextRank algorithm based on upper approximation rough data-deduction. The experimental results show that the accurate rate, recall rate, and $F1$ of proposed algorithm have been significantly improved, and the experimental results also demonstrate that the proposed algorithm has good performance in running time and physical memory occupation.

1. Introduction

The mental health and wellbeing of undergraduates have deteriorated over the last decade. Before the COVID-19 pandemic, higher education was facing a "mental health crisis" [1, 2]. The rapid onset of the COVID-19 pandemic has introduced countless additional stressors, and faculty concern over student wellbeing has increased. Over the past ten or twenty years, the depression has increased from about 25% of undergraduates in 2010 to almost 30% of undergraduates in 2020, and the anxiety of undergraduates has increased from 22% in 2014 to 31% in 2020. Suicidal ideation of undergraduates has increased from 6% in 2010 to 11% in 2020 [3]. The frequency of mental health management organization in undergraduates varies from university to university. Definitely influence of the pandemic on mental health concerns within undergraduates is a big concern. The pandemic has affected the economic development of many countries, and the cooperation of relevant countries on the pandemic has also led to conflicts. The widespread public reports on the Internet and the media have made simple and inexperienced undergraduates unable to distinguish. So, the management and plan of undergraduates' mental health are important under COVID-19 pandemic [4, 5].

Keywords are words that express the central content of a document. Keywords from a document can accurately describe the document's content and facilitate fast information processing. There are two main types of keyword extraction algorithms, which are unsupervised keyword extraction method and supervised keyword extraction method [6–8]. Unsupervised keyword extraction method does not need manually labeled corpus, but it uses some methods to find important words in the text as keywords for keyword extraction. In unsupervised keyword extraction method, candidate words are firstly extracted, and then each candidate word is scored, so top-K candidate words with the highest score are output as keywords. According to different ranking strategies, there are different algorithms such as term frequency-inverse document frequency (TF-IDF), TextRank, and latent Dirichlet allocation (LDA). The supervised keyword extraction method regards the keyword extraction process as a binary classification problem. At first, the candidate words are extracted, and then each candidate word is labeled, so the keyword extraction classifier is

trained. When a new document is coming, all candidate words are extracted, and then the trained keyword extraction classifier is used to classify each candidate word. Finally, the candidate words labeled as keywords are used as keywords [9, 10].

Accordingly, the main contributions of this paper are summarized as follows. (i) I study the TextRank keyword extraction algorithm. (ii) An improved TextRank algorithm based on upper approximation rough data-deduction is proposed.

The rest of this paper is structured as follows. Section 2 reviews the related work. In Section 3, I propose an improved TextRank algorithm based on upper approximation rough data-deduction. The experimental results are shown in Section 4. Section 5 concludes this paper.

2. Related Work

Many strategies of management and plan for undergraduates' mental health have been proposed. In [11], the authors studied to examine student perspectives about college mental health including the primary mental health issues affecting students, common college student stressors, student awareness of campus mental health resources, and mental health topics students wanted more information about. Little research existed into the trends associated with on-campus service utilization for mental health concerns of college students. Rates of broad service utilization existed, but no published study had examined the direct relationship between a range of common mental health symptoms and on-campus service utilization. In [12], the authors studied to explore which common mental health concerns were associated with specific on-campus service utilization in undergraduate students and whether endorsement of more mental health concerns would predict a higher number of services utilized. In [13], the study investigated the moderating role of perceived social support in the relationship between academic demands (measured as perceived academic stress) and mental health of undergraduate students in full-time employment. A growing number of developing countries had experienced worsening air pollution, which had been shown to cause significant health problems. However, few studies had explored the impact of air pollution on the mental health of university students, particularly in the Chinese context. In order to address this gap, in [14], through a large-scale cross-sectional survey, the study aimed to examine the effects of air pollution on final-year Chinese university undergraduates' mental health by employing multivariable logistic regression.

The TextRank algorithm plays an important role in keyword extraction. In [15], the author presented an automatic keyword extraction algorithm based primarily on a weighted TextRank model. In the model, word embedding vectors were used to compute a similarity measure as an edge weight. As a typical keyword extraction technology, TextRank had been used in a wide variety of commercial applications, including text classification, information retrieval, and clustering. In these applications, the parameters of TextRank, including the cooccurrence window size, iteration number, and decay factor, were set roughly. In [16], the authors conducted an empirical study on TextRank, towards finding optimal parameter settings for keyword extraction. The keyword weight propagation in TextRank focused only on word frequency. To improve the performance of the algorithm, in [17], the authors proposed semantic clustering TextRank, a semantic clustering news keyword extraction algorithm based on TextRank. In [18], the authors introduced a new human-annotated Chinese patent dataset and proposed a sentence-ranking-based term frequency-inverse document frequency algorithm for patent keyword extraction, motivated by the thought of "the keywords were in the key sentences." In the algorithm, a sentence-ranking model was constructed to filter top-K-s percent sentences from each patent based on a sentence semantic graph and heuristic rules. In [19], the authors introduced a word network whose nodes represented words in a document and defined that any keyword extraction method based on a word network was called as a Word-net method. Then, the authors proposed a new network model which considered the influence of sentences and a new word-sentence method based on the new model. In [20], the authors proposed an ontology and enhanced word embedding-based methodology for automatic keyphrase extraction from geoscience documents.

There are also some other methods for keyword extraction. In [21], an enhancement of the term weighting was proposed particularly in the form of a series of modified term frequency-inverse document frequencies, for improving keyword extraction. In [22], the authors proposed an improved rapid automatic keyword extraction method, which used the word string matching feature in the dictionary method to correspond to the relevant execution action function. In [23], a novel text mining approach based on keyword extraction and topic modeling was introduced to identify key concerns and their dynamics of on-site issues for better decision-making process.

3. Improved Algorithm Based on Rough Data-Deduction

TextRank keyword extraction algorithm is a graph-based ranking algorithm, which is derived from Google's PageRank algorithm [24]. TextRank firstly divides the target text into several meaningful words and constructs the candidate word graph and then uses the voting mechanism to rank the candidate words to achieve keyword extraction. The task of keyword extraction is to extract several important words from the target text. TextRank algorithm uses the local correlation between words (i.e., cooccurrence sliding window) to determine the correlation between candidate words and then performs iterative calculation and ranking of candidate keywords. Rough set theory is originally used for text classification to speed up classification and improve accuracy. Rough data-deduction is based on rough set theory, which integrates approximate information from upper approximation concept into data reasoning process. This paper introduces upper approximation-based rough data-deduction to TextRank keyword extraction algorithm,

and the extracted keywords are used in undergraduates' mental health management and plan.

In TextRank keyword extraction algorithm, the candidate keywords in the text are the graph model constructed by the cooccurrence correlation, and then the weights of each node are calculated by the average transition probability matrix for many times until convergence. After convergence, words are ranked in descending order according to their weights, and the first N words are selected as the extracted keywords. This method is more concise and effective, but it has certain limitations. In [25], the convergent operation utilizes a clustering strategy to group the population into multiple clusters. The use of cooccurrence window only considers the correlation between local words, so some words closely related to a certain keyword may be ignored, but keywords from a document are not just limited to the keywords around the words. When doing text keyword extraction, I should fully consider the words in the text as well as some potentially related words. Words with potential correlation will have an important impact on the whole iterative ranking process, and the potential relation can be discovered by the theory of rough data-deduction. Therefore, this paper proposes an improved TextRank algorithm based on upper approximation rough data-deduction.

Based on the word sense similarity of mental health words, the candidate keywords are divided. As there may be a group of words with similar word sense in a document, the weight of this group of words should be increased to improve the accuracy of extraction results when describing the same important content. TextRank algorithm only considers the word sense themselves and ignores the contribution of words with similar word sense. Therefore, the improved algorithm takes the word sense into account and divides the candidate words by word sense, which can extract keywords more effectively.

The rough data-deduction space $M = (U, N, D)$ is introduced to describe the keyword extraction of undergraduates' mental health issue structurally. U is the universe of discourse (UOD) and the dataset composed of candidate keywords of undergraduates' mental health. N is a set of equivalence relation, and $E \in N$. If and only if p is similar to q, then $p, q \in U$ and $<p, q> \in E$. $D \subseteq U \times U$ is defined as $D = \{<p, q> | p, q \in U$ and there is a relation between p and $q\}$.

Assuming that deduction correlation is defined as equation (1) by using rough data-deduction,

$$D = \{\langle cw_1, cw_4\rangle, \langle cw_2, cw_6\rangle, \langle cw_3, cw_6\rangle, \langle cw_6, cw_5\rangle\}, \tag{1}$$

where cw_1–cw_6 are the candidate keywords from the text through word segmentation and filtering, and the deduction correlation is determined by the degree of the association rules, that is, pointwise mutual information (PMI).

At the same time, for equivalence relation $E \in N$,

$$\frac{U}{E} = \{\{cw_1, cw_2, cw_3\}, \{cw_4, cw_6\}, \{cw_5, cw_7\}\}, \tag{2}$$

where the equivalence division is based on the similarity between the candidate words.

In rough data-deduction, for candidate word cw_1, the algorithm obtains cw_2 and cw_3 based on similarity rule and then divides cw_1, cw_2, and cw_3 into one dataset, and cw_4–cw_7 can be similarly divided. Then, cw_4 can be obtained from cw_1 based on the degree of the association rules of PMI, as well as cw_5, cw_6, and cw_7. According to rough data-deduction, for cw_1, $[cw_1]_E = \{cw_1, cw_2, cw_3\}$, and $[cw_1 - E] = \{cw_4, cw_6\}$. $E^*([cw_1 - E]) = \{cw_4, cw_6\}$, so $cw_1 \Rightarrow_E cw_6$. For candidate word cw_6, $[cw_6]_E = \{cw_4, cw_6\}$, and $[cw_6\text{-}E] = \{cw_5\}$. $E^*([cw_6 - E]) = \{cw_5, cw_7\}$, so $cw_6 \Rightarrow_E cw_7$. $Cw_1 =_E cw_7$ can be obtained from $cw_1 \Rightarrow_E cw_6$ and $cw_6 \Rightarrow_E cw_7$. As described, there is also a potential correlation between cw_1 and cw_7, which can provide a certain contribution rate for calculation. The association between candidate keywords is established by the above rules, and the association weight can be added to the iterative calculation process as contribution rate to improve the accuracy of keyword extraction.

The upper approximation-based rough data-deduction to TextRank keyword extraction algorithm is summarized as follows.

Step 1. Based on TextRank algorithm, the text related to undergraduates' mental health is preprocessed, which includes clause, word segmentation, and part of speech (POS) tagging, and candidate keywords are obtained.

Step 2. The candidate keywords are divided into different equivalence classes according to their similarities. This paper is divided based on WordNet and Wikitext. For any two candidate words cw_1 and cw_2, the division rule is defined as follows:

$$s = \omega_1 s_1 + \omega_2 s_2, \tag{3}$$

where s_1 and s_2 are the similarities calculated by WordNet and Wikitext, respectively. ω_1 and ω_2 are the two weights assigned to s_1 and s_2, and $\omega_1 + \omega_2 = 1$.

Assuming that candidate word cw_1 is distributed in WordNet WN, and cw_2 is distributed in Wikitext WT, the intersection of WN and WT is WW. The value strategy of ω_1 and ω_2 is summarized as follows:

(i) When $cw_1 \in WW$ and $cw_2 \in WW$, the similarity to cw_1 and cw_2 is calculated based on WN and WT, respectively, which are denoted as s_1 and s_2. In this paper, $\omega_1 = \omega_2 = 0.5$.

(ii) When $cw_1 \in WN$ and $cw_2 \in WN$, or $cw_1 \in WT$ and $cw_2 \in WT$, cw_1 and cw_2 are calculated as s_1 or s_2 based on WN and WT, where one of the ω_1 and ω_2 is 1, and the other one is 0.

(iii) When $cw_1 \in WN$ and $cw_2 \in WT$, the synonym set of cw_2 is searched based on WT, and then the similarity with cw_1 is calculated based on WN, and the maximum value is denoted as s_1. If cw_2 has no synonym in WT, then $s_1 = 0.2$, $\omega_1 = 1$, and $\omega_2 = 0$.

(iv) When $cw_1 \in WN$ and $cw_2 \in WW$, the similarity to cw_1 and cw_2 is calculated based on WN and denoted as s_1. Then, the synonym set of cw_2 is searched in WT, and then the similarity to cw_1 is calculated based on

WN, and the maximum value is denoted as s_2. If cw_2 has no synonym in *WT*, then $s_2 = s_1$, and $\omega_1 > \omega_2$. In this paper, $\omega_1 = 0.6$, and $\omega_2 = 0.4$.

(v) When $cw_1 \in WT$ and $cw_2 \in WW$, the similarity to cw_1 and cw_2 is calculated based on *WT* and denoted as s_2. Then, the synonym set of cw_1 is searched in *WT*, and then the similarity to cw_2 is calculated based on *WN*, and the maximum value is denoted as s_1. If cw_1 has no synonym in *WT*, then $s_1 = s_2$, and $\omega_2 > \omega_1$. In this paper, $\omega_1 = 0.4$, and $\omega_2 = 0.6$.

Here, the calculation of word similarity based on WordNet is defined as follows:

$$s(WW_1, WW) = \sum_{m=1}^{3} \ln \delta_m \prod_{n=1}^{m} s_n(WW_1, WW_2), \quad (4)$$

$$s(cw_1, cw_2) = \max_{m=1\ldots i, n=1\ldots j} s(WW_{1m}, WW_{2n}). \quad (5)$$

In equation (4), $s_1(WW_1, WW_2)$ is the similarity calculated by the set of independent minimum semantic units. $s_2(WW_1, WW_2)$ is the similarity of feature structure of minimal semantic unit of correlation. $s_3(WW_1, WW_2)$ is the similarity of the characteristic structure of the relational sign. The parameter δ_m $(1 \le m \le 3)$ is adjustable and meets the requirement of $\delta_1 + \delta_2 + \delta_3 = 1$. In this paper, δ_1, δ_2, and δ_3 are set as 0.6, 0.25, and 0.15, respectively. Equation (5) can obtain the sense similarity. When there are multiple senses in a word, equation (5) is used to calculate the maximum similarity among all combinations of senses, that is, the similarity to two words, where i is the sense number of the word cw_1, and j is the sense number of the word cw_2.

The calculation of word similarity based on *WT* is defined as follows:

$$s(WW_1, WW_2) = (1 - 0.5\mathrm{d}t(WW_1, WW_2))\sqrt{e^{-\mathrm{d}i/2j}}, \quad (6)$$

where $dt(WW_1, WW_2)$ is the distance function of word codes WW_1 and WW_2 in the tree structure. j is the total number of nodes in the branch layer, which indicates the number of direct child nodes of the nearest common parent node of two words. di represents the distance between branches where two words are located in the nearest public parent node.

Step 3. The correlation of association rules in rough data-deduction is defined as follows:

$$\mathrm{PMI}(cw_1, cw_2) = \frac{p(CW_1, CW_2)}{p(CW_1)p(CW_2)}, \quad (7)$$

where cw_1 and cw_2 are two candidate keywords in the text. $p(cw_1, cw_2)$ is the probability of cw_1 and cw_2 appearing in the same sentence. $p(cw_1)$ is the probability of occurrence of cw_1, and $p(cw_2)$ is the probability of occurrence of cw_2.

According to the correlation, the candidate keywords with direct correlation are determined, when $PMI(cw_1, cw_2) \neq 0$, there is a direct correlation between cw_1 and cw_2, and cw_1, cw_2 and their correlation degrees are stored in the correlation set. Meanwhile, the rough data-deduction relation *D* can be established according to the correlation degree.

Then, by using the rules of rough data-deduction, I get the correlation between the other candidate keywords in all the different equivalence classes, and these words and their correlation degrees are stored into the correlation set.

Step 4. According to the correlation set obtained in Step 3, candidate keyword graphs with weights are constructed. Then, according to the equation of TextRank algorithm, the weight of each candidate keyword is calculated iteratively until convergence.

4. Experiment and Results Analysis

4.1. Experimental Data and Evaluation Criteria. The experiment selects 26300 questionnaires of mental health management of undergraduates from Xinxiang University with 23 schools and 60 majors, which consist of psychological distress, depression, suicidal tendency, and self-evaluation related to mental health within 300 to 1000 words. In particular, these undergraduates are distributed for different grades uniformly. The 19000 valid questionnaires are obtained by excluding questionnaires with self-evaluation less than 300 words to test the effect of proposed method in this paper. The questionnaires use silver ink with a metal oxide [26]. 10 keywords of each questionnaire are extracted and ranked by the importance. In this paper, ω_1 and ω_2 are set 0.5.

In addition, for comparison purposes, TextRank, a keyword extraction using supervised cumulative TextRank (KESCT) [27], scientific research project TF-IDF (SRP-TF-IDF) [28], and high representation tags LDA (HRT-LDA) [29] are selected. Three evaluation indexes commonly used in classification are used to compare and evaluate the quality of experimental results, which include precision (*P*), recall rate (*R*), and *F*1 (*F*). *P* is the accuracy of extraction results. *R* is the coverage degree of the extraction results to the correct keywords. *F* is a comprehensive evaluation index of harmonic average of *P* and *R*.

4.2. Experimental Results. It is found in the experiment that the two important parameters can affect the keyword extraction result of TextRank algorithm, which are the cooccurrence window size and the number of keywords, while the implementation of TF-IDF algorithm based on statistical feature and the algorithm proposed in this paper are not affected by the cooccurrence window size. I set the number of extracted keywords as 10, and the value of the comparison window is within [4, 10]. The *F*1 under different cooccurrence window sizes is shown in Figure 1.

It can be seen from Figure 1 that TextRank algorithm has different extraction effects under different cooccurrence window sizes. In the same test set, this paper compares the effect of different cooccurrence window sizes, and when the cooccurrence window size is 5, the original TextRank algorithm has the best extraction effect with high *F* value.

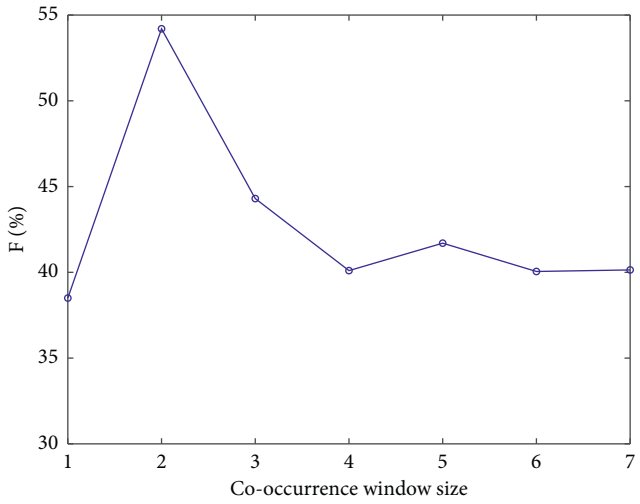

Figure 1: F value under different cooccurrence window sizes.

Therefore, in order to ensure the effectiveness of the proposed algorithm, the cooccurrence window size is set to 5.

The initial window value is set to 5, and P, R, and F are calculated with the number of keywords within [3, 10]. The calculation results are shown in Table 1.

At the same time, in order to observe the experimental results of five algorithms conveniently, the P, R, and F of the algorithm are plotted, as shown in Figures 2–4.

Figure 2 describes the variation trend of the accuracy of the five algorithms when extracting different numbers of mental health keywords. As can be seen from Figure 2, with the increasing number of mental health keywords extracted, the accuracy of each algorithm decreases, but the accuracy of the algorithm proposed in this paper is always higher than other four baselines. The TextRank algorithm based on rough data-deduction proposed in this paper will integrate upper approximation information into the process of data-deduction so that the mutual deduction between data presents the characteristics of approximate entailment or imprecise association, and the potential association between candidate keywords can be mined. If the potential association is added to the iterative calculation of the weight of each candidate keyword, more accurate extraction results can be obtained. Therefore, the accuracy of the algorithm proposed in this paper is theoretically higher, and its accuracy P value is higher than other four baselines.

Figure 3 describes the change of recall rate of five algorithms when extracting different numbers of mental health keywords. In Figure 3, the recall rate of the algorithm proposed in this paper is higher than that of other four baselines, and the recall rate increases with the increasing number of mental health keywords. The SRP-TF-IDF algorithm relies too heavily on word frequency and does not use correlation between words at all. KESCT algorithm adopts the cooccurrence window principle. Although the relation between words is considered, the algorithm is more inclined to put forward frequent words due to its limitations, which may ignore important words with low word frequency that can describe the topics. However, the rough data-deduction used in this paper can expand the correlation range and enhance the coverage of the keywords of the correct correlation in order to improve the recall rate of the algorithm. The influence of word frequency decreases with the increasing number of keywords, and the advantages of the algorithm proposed in this paper will be more obvious.

Figure 4 describes the F values of five algorithms when extracting different numbers of mental health keywords. When evaluating the experimental results, it is expected that both P and R should be as high as possible. However, in most cases, the two values are contradictory. Therefore, F value should be used to comprehensively consider the two values, which can reflect the effectiveness of the whole algorithm. Keyword extraction based on rough data-deduction can mine the potential association between candidate keywords theoretically, which increases the candidate words and range of the association. The keyword extraction based on rough data-deduction adds the potential association to the iterative

Table. 1: The comparison of experimental results of five algorithms.

The number of keywords	Algorithm	P (%)	R (%)	F (%)
	TextRank	52.32	15.50	23.79
	KESCT	55.03	22.35	30.08
3	SRP-TF-IDF	64.10	23.00	33.59
	HRT-LDA	70.54	24.56	34.16
	This paper	80.67	29.30	36.51
	TextRank	49.06	19.71	28.01
	KESCT	54.76	24.33	33.08
4	SRP-TF-IDF	62.15	26.94	34.15
	HRT-LDA	69.87	29.41	35.09
	This paper	79.06	31.08	43.66
	TextRank	48.51	25.16	32.18
	KESCT	52.61	27.66	36.07
5	SRP-TF-IDF	60.48	28.54	37.22
	HRT-LDA	66.92	30.69	39.05
	This paper	78.35	32.17	47.65
	TextRank	42.67	27.65	32.01
	KESCT	49.62	30.25	36.95
6	SRP-TF-IDF	55.40	32.49	34.12
	HRT-LDA	60.73	34.16	40.08
	This paper	77.70	39.58	51.89
	TextRank	41.19	29.90	33.77
	KESCT	45.12	34.15	33.99
7	SRP-TF-IDF	50.08	36.58	40.06
	HRT-LDA	55.64	38.04	45.37
	This paper	74.39	45.89	54.68
	TextRank	39.66	31.57	34.55
	KESCT	40.51	35.26	32.89
8	SRP-TF-IDF	45.99	38.16	39.98
	HRT-LDA	49.30	42.53	44.38
	This paper	70.88	50.07	55.81
	TextRank	36.78	32.36	33.99
	KESCT	37.41	36.45	34.01
9	SRP-TF-IDF	44.01	40.19	40.71
	HRT-LDA	40.09	45.20	46.39
	This paper	65.72	55.06	56.09
	TextRank	35.87	35.64	33.87
	KESCT	33.02	40.16	33.66
10	SRP-TF-IDF	38.28	42.26	38.74
	HRT-LDA	39.16	49.39	43.91
	This paper	60.10	59.03	55.99

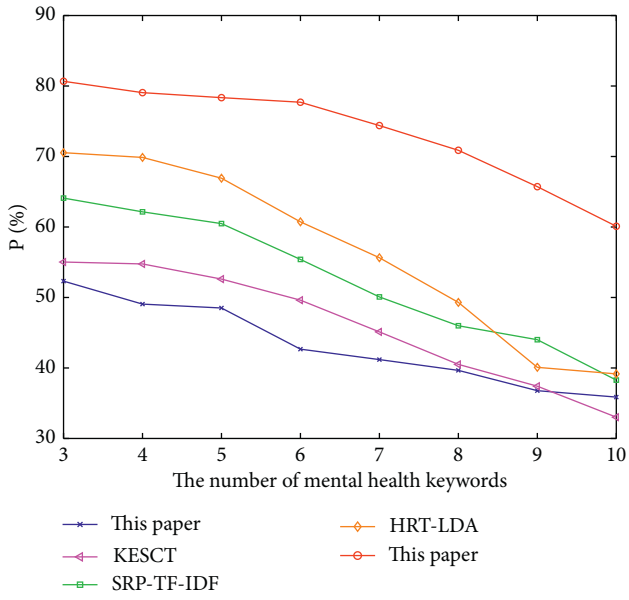

Figure 2: The comparison of P of the five algorithms.

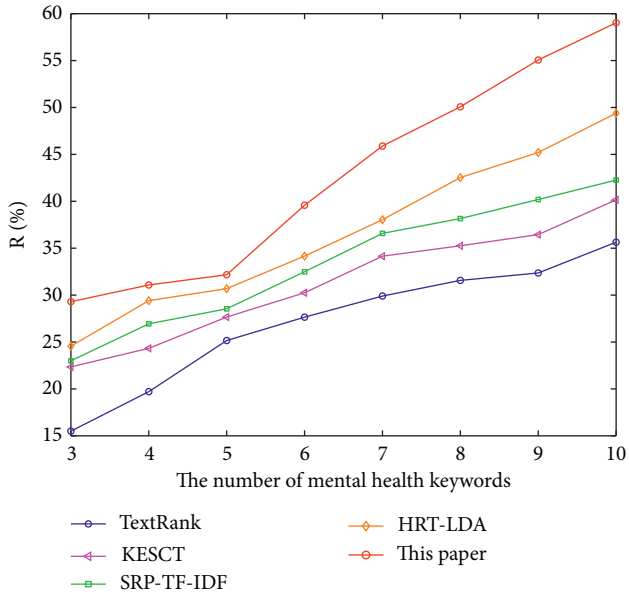

Figure 3: The comparison of R of the five algorithms.

calculation of the weight of each candidate keyword, so the extraction results will be more accurate, that is, the algorithm is also more effective.

According to the experimental results, the proposed algorithm has higher P and R than the other four baselines. The F will be higher with the higher P and R, and the higher F can indicate the effectiveness of the algorithm. In conclusion, the accuracy rate, recall rate, and comprehensive evaluation index $F1$ of the proposed algorithm are higher than those of the four baselines, which indicates that the improved TextRank algorithm based on upper approximation rough data-deduction is more effective in mental health keyword extraction.

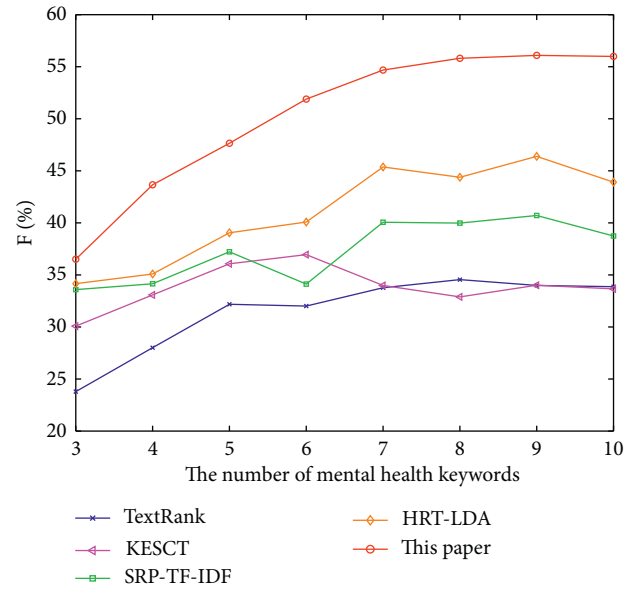

Figure 4: The comparison of F of the five algorithms.

The test set of this paper uses the self-evaluation in the questionnaire of undergraduates' mental health. The text length is generally less than 500 words, which is mainly concentrated in 300–500 words. This paper divides the test set by the number of self-evaluation words. Each test set randomly selects 30 texts of corresponding text words to compare the running time and physical memory occupation of the five algorithms.

As can be seen from Figure 5, when the number of words in the text is 300–400, the number of deduction and semantic calculation is small, and the running time of the algorithm is also short. The number of deduction and semantic calculations increases with the increasing number of words. Compared with TextRank, the running time of the proposed method is still shorter than that of the other three baselines, which is similar to TextRank's efficiency.

It can be seen from Figure 6 that the physical memory occupation of the proposed method is small with good efficiency. When the number of words in the text is 600–800 or 800–1000, the physical memory occupation of SRP-TF-IDF and KESCT is nearly the same.

This paper manages the undergraduates' mental health through the keyword extraction. The results show that academic problem, emotional problem, interpersonal problem, anxiety problem, sexual problem, and adaptation to college life are the universal mental health issues of undergraduates. Currently, how to deal with mental crisis is an urgent problem that colleges cannot avoid. The proposed bounded area elimination algorithm in [30] analyzes the feature extraction, and the idea of feature extraction is similar to the TextRank keyword extraction algorithm proposed in this paper . Timely plan of mental health crisis is to provide supports and help to those who have experienced personal crisis so that they can restore their mental balance and have full confidence in life.

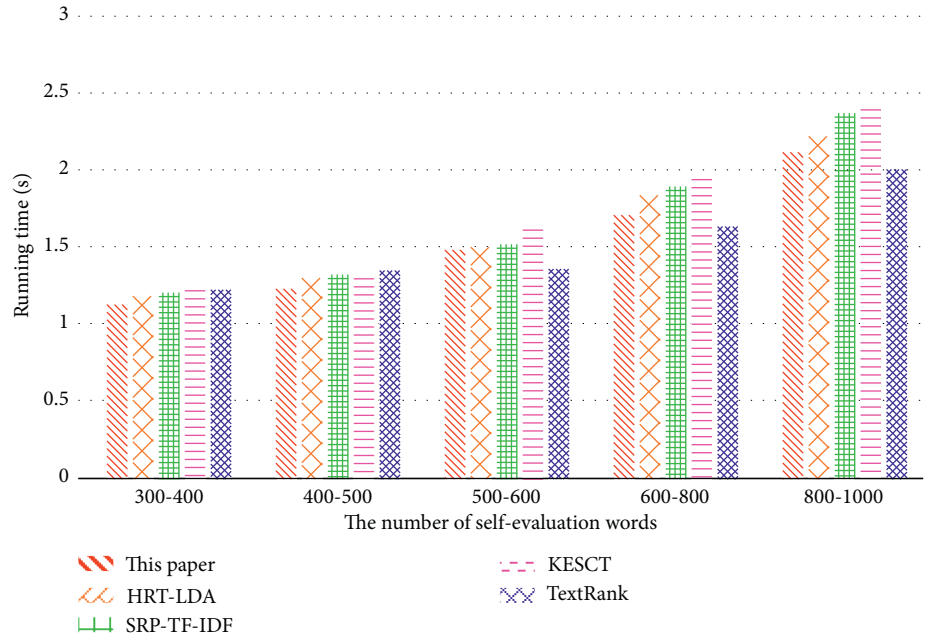

Figure 5: The comparison of running time of keyword extraction algorithms.

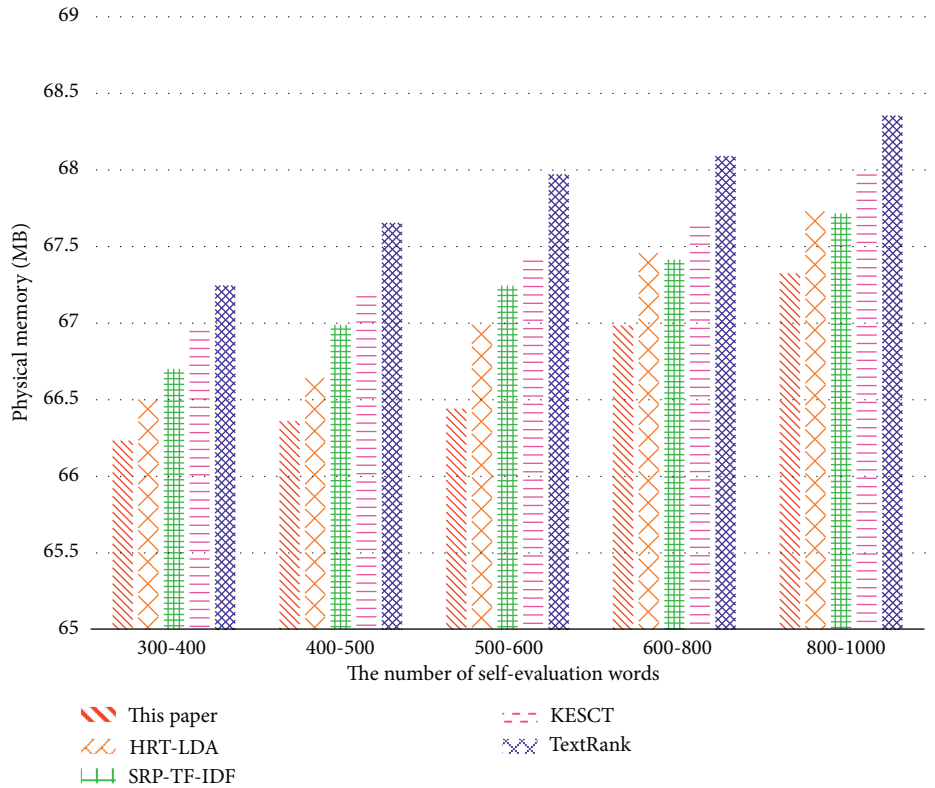

Figure 6: The comparison of physical memory occupation of keyword extraction algorithms.

5. Conclusions

In a fast-paced society, there is more competition among undergraduates, that is, they are facing the dual pressure of enrollment and employment, and mental health is very important for them. Mental health is the necessary condition and foundation for everyone's all-round development in today's society, which is also a necessary psychological quality for undergraduates. This paper introduces upper approximation-based rough data-deduction to TextRank

keyword extraction algorithm, and the extracted keywords are used in undergraduates' mental health management and plan. The comparison experiments reveal that the proposed algorithm outperforms four baselines in terms of accuracy rate, recall rate, $F1$, running time, and physical memory occupation.

The future works are stated as follows. (i) The deduction rules of rough data will be further refined and improved, so as to get better extraction effect. (ii) The words related to mental health may be incomplete in WordNet and Wikitext, which results in unsatisfactory keyword extraction. The following research will consider using a corpus of mental health related to achieve keyword extraction.

Data Availability

All data used to support the findings of the study are included within the article.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] S. Rosenberg, J. Mendoza, H. Tabatabaei-Jafari, and L. Salvador-Carulla, "International experiences of the active period of COVID-19 - mental health care," *Health Policy and Technology*, vol. 9, no. 4, pp. 503–509, 2020.

[2] S. Johnson, C. Dalton-Locke, N. Vera San Juan et al., "Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff," *Social Psychiatry and Psychiatric Epidemiology*, vol. 56, no. 1, pp. 25–37, 2021.

[3] J. D. Buckner, E. M. Lewis, and R. P. Tucker, "Mental health problems and suicide risk: the impact of acute suicidal affective disturbance," *Archives of Suicide Research*, vol. 24, no. sup1, pp. 303–313, 2020.

[4] M. M. Kibbey, E. J. Fedorenko, and S. G. Farris, "Anxiety, depression, and health anxiety in undergraduate students living in initial US outbreak "hotspot" during COVID-19 pandemic," *Cognitive Behaviour Therapy*, vol. 50, no. 5, pp. 409–421, 2021.

[5] A. Saha, A. Dutta, and R. I. Sifat, "The mental impact of digital divide due to COVID-19 pandemic induced emergency online learning at undergraduate level: evidence from undergraduate students from Dhaka City," *Journal of Affective Disorders*, vol. 294, pp. 170–179, 2021.

[6] N. Firoozeh, A. Nazarenko, F. Alizon, and B Daille, "Keyword extraction: issues and methods," *Natural Language Engineering*, vol. 26, no. 3, pp. 259–291, 2020.

[7] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, "Yake! keyword extraction from single documents using multiple local features," *Information Sciences*, vol. 509, pp. 257–289, 2020.

[8] S. Duari and V. Bhatnagar, "sCAKE: semantic connectivity aware keyword extraction," *Information Sciences*, vol. 477, pp. 100–117, 2019.

[9] X. Mao, S. Huang, R. Li, and L. Shen, "Automatic keywords extraction based on co-occurrence and semantic relationships between words," *IEEE Access*, vol. 8, pp. 117528–117538, 2020.

[10] A. K. Bhunia, P. P. Roy, A. Sain, and U. Pal, "Zone-based keyword spotting in Bangla and Devanagari documents," *Multimedia Tools and Applications*, vol. 79, no. 37-38, pp. 27365–27389, 2020.

[11] S. Gibbons, T. Trette-McLean, A. Crandall, J. L. Bingham, C. L. Garn, and J. C. Cox, "Undergraduate students survey their peers on mental health: perspectives and strategies for improving college counseling center outreach," *Journal of American College Health*, vol. 67, no. 6, pp. 580–591, 2019.

[12] J. L. Bourdon, A. A. Moore, E. C. Long, K. S. Kendler, and D. M. Dick, "The relationship between on-campus service utilization and common mental health concerns in undergraduate college students," *Psychological Services*, vol. 17, no. 1, pp. 118–126, 2020.

[13] C. U. Onuoha and E. S. Idemudia, "Academic demands and mental health among undergraduate students in full-time employment: the moderating role of perceived social support," *Journal of Psychology in Africa*, vol. 30, no. 2, pp. 89–95, 2020.

[14] D. Zu, K. Zhai, Y. Qiu, P. Pei, X. Zhu, and D. Han, "The impacts of air pollution on mental health: evidence from the Chinese university students," *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, 2020.

[15] P. Wongchaisuwat, "Automatic keyword extraction using textrank," in *Proceedings of the IEEE 6TH International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 377–381, Tokyo Japan, April 2019.

[16] M. Zhang, X. Li, S. Yue, and L. Yang, "An empirical study of TextRank for keyword extraction," *IEEE Access*, vol. 8, pp. 178849–178858, 2020.

[17] A. Xiong, D. Liu, H. Tian, Z. Liu, P. Yu, and M. Kadoch, "News keyword extraction algorithm based on semantic clustering and word graph model," *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 886–893, 2021.

[18] Z. Wang and Y. Guo, "Sentence-ranking-enhanced keywords extraction from Chinese patents," *Journal of Information Science and Engineering*, vol. 35, no. 3, pp. 651–674, 2019.

[19] L. Yang, K. Li, and H. Huang, "A new network model for extracting text keywords," *Scientometrics*, vol. 116, no. 1, pp. 339–361, 2018.

[20] Q. Qiu, Z. Xie, L. Wu, and W Li, "Geoscience keyphrase extraction algorithm using enhanced word embedding," *Expert Systems with Applications*, vol. 125, pp. 157–169, 2019.

[21] B. Chiraratanasopha, S. Boonbrahm, and T. Theeramunkong, "Effect of term weighting on keyword extraction in hierarchical category structure," *Computing and Informatics*, vol. 40, no. 1, pp. 57–82, 2021.

[22] C.-H. Shih, C.-J. Lin, and S.-Y. Jeng, "Improved rapid automatic keyword extraction for voice-based mechanical arm control," *Sensors and Materials*, vol. 33, no. 8, pp. 2897–2909, 2021.

[23] J.-R. Lin, Z.-Z. Hu, J.-L. Li, and L-M Chen, "Understanding on-site inspection of construction projects based on keyword extraction and topic modeling," *IEEE Access*, vol. 8, pp. 198503–198517, 2020.

[24] P. Carmona and J. L. Castro, "fuzzyfeaturerank. bringing order into fuzzy classifiers through fuzzy expressions," *Fuzzy Sets and Systems*, vol. 401, pp. 78–90, 2020.

[25] L. Ma, S. Cheng, and Y. Shi, "Enhancing learning efficiency of brain storm optimization via orthogonal learning design," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 11, pp. 6723–6742, 2021.

[26] L. Yung, C. Fei, J. Mandeep, H. Abdullah, and L. Wee, "Synthesis of a nano-silver metal ink for use in thick conductive film fabrication applied on a semiconductor package," *PLoS One*, vol. 9, no. 5, 2014.

[27] M. Bordoloi, P. C. Chatterjee, S. K. Biswas, and B Purkayastha, "Keyword extraction using supervised cumulative textrank," *Multimedia Tools and Applications*, vol. 79, no. 41-42, pp. 31467–31496, 2020.

[28] Z. Wang, D. Wang, and Q. Li, "Keyword extraction from scientific research projects based on SRP-TF-IDF," *Chinese Journal of Electronics*, vol. 30, no. 4, pp. 652–657, 2021.

[29] Y. Zhao, Y. Qiao, and K. He, "A novel tagging augmented LDA model for clustering," *International Journal of Web Services Research*, vol. 16, no. 3, pp. 59–77, 2019.

[30] H. Y. Chai, L. K. Wee, T. T. Swee, S. H. Salleh, and L. Y. Chea, "An artifacts removal post-processing for epiphyseal region-of-interest (EROI) localization in automated bone age assessment (BAA)," *BioMedical Engineering Online*, vol. 10, p. 87, 2011.