Scientometrics (2022) 127:2803–2827
https://doi.org/10.1007/s11192-022-04350-5

Measuring the innovation of method knowledge elements in scientific literature

Zhongyi Wang[1] · Keying Wang[1] · Jiyue Liu[1] · Jing Huang[2] · Haihua Chen[3]

Received: 3 October 2021 / Accepted: 8 March 2022 / Published online: 25 March 2022

Abstract

Interest in assessing research impacts is increasing due to its importance for informing actions and funding allocation decisions. The level of innovation (also called "innovation degree" in the following article), one of the most essential factors that affect scientific literature's impact, has also received increasing attention. However, current studies mainly focus on the overall innovation degree of scientific literature at the macro level, while ignoring the innovation degree of a specific knowledge element (KE), such as the method knowledge element (MKE). A macro level view causes difficulties in identifying which part of the scientific literature contains the innovations. To bridge this gap, a more fine-grained evaluation of academic papers is urgent. The fine-grained evaluation method can ensure the quality of a paper before being published and identify useful knowledge content in a paper for academic users. Different KEs can be used to perform the fine-grained evaluation. However, MKEs are usually considered as one of the most essential knowledge elements among all KEs. Therefore, this study proposes a framework to **m**easure the **inno**vation **d**egree of **m**ethod **k**nowledge **e**lements (**MIDMKE**) in scientific literature. In this framework, we first extract the MKEs using a rule-based approach and generate a cloud drop for each MKE using the biterm topic model (BTM). The generated cloud drop is then used to create a method knowledge cloud (MKC) for each MKE. Finally, we calculate the innovation score of a MKE based on the similarity between it and other MKEs of its type. Our empirical study on a China National Knowledge Infrastructure (CNKI) academic literature dataset shows the proposed approach can measure the innovation of MKEs in scientific literature effectively. Our proposed method is useful for both reviewers and funding agencies to assess the quality of academic papers. The dataset, the code for implementation the algorithms, and the complete experiment results will be released at: https://github.com/haihua0913/midmke.

Keywords Academic literature · Method knowledge element · Biterm topic model · Cloud model · Similarity cloud · Innovation degree

✉ Haihua Chen
haihua.chen@unt.edu

[1] School of Information Management, Central China Normal University, Wuhan 430079, China

[2] Wuhan Polytechnic, Wuhan 430023, China

[3] Department of Information Science, University of North Texas, Denton, TX 76203, USA

Introduction

Scientific literature, describing new ideas and consolidating existing concepts in the sciences, is essential to understanding advances in different fields. Measuring the innovation in scientific literature is vital from both a management and a policy standpoint (Funk & Owen-Smith, 2017). Innovation in scientific literature directly reflects the quality and potential value of both research and good researchers. Nevertheless, measuring the quality, especially the level of innovation in scientific literature, could be a valuable capability, though intractable.

Recently, there is growing interest in measuring the innovation degree of academic literature. Scholars have used various methods for measuring innovation in scientific literature. Approaches used can be summarized into two categories: metadata-based bibliometric measures and content-based measures.

Metadata-based bibliometric measures study innovation in academic literature mainly through impact assessment. These measures can take different forms, such as correlating innovation with authors and sponsors, or taking citations and keywords as an indicator to evaluate the quality of scientific literature. Examples of the former are methods to appraise scientific literature according to the levels of authors (Tahamtan et al., 2016) and sponsors (Wang et al., 2018), while the latter include the journal impact factor (Garfield, 2006), h-index (Hirsch, 2005), and so on. Metadata-based bibliometric measures are efficient and have the potential to detect scientific literature with high innovation automatically, but the impact assessment of any scientific literature only reflects what happens after publication.

Compared to metadata-based bibliometric measures, content-based measures can reduce the time lag by directly measuring the innovation expressed by the text data of scientific literature. These measures use text mining techniques that scan a large volume of textual data to identify the degree of innovation in academic papers. For example, some researchers tracked the frequency of topics over time and used the frequency score to indicate innovation of topics (Mörchen et al., 2008). Some studies employed a topic's temporal relationship with other topics to decide the innovation of topics (He et al., 2009; Yan, 2014). These techniques can automatically detect research topics based on textual information and identify their innovation.

However, to the best of our knowledge, there are as yet no effective methods since most of them introduce different biases during the evaluation. First, recent work has shown that metadata-based bibliometric measures like citation counts are biased against newer articles Wang et al. (2017). Second, content-based measures mainly focus on the overall innovation degree of scientific literature at the macro level (Reich, 1995; He & Chen, 2018). They pay little attention to the innovation degree of specific knowledge claims in scientific literature at the micro level of knowledge elements (KEs). Ding et al. (2013) argued that looking deeper into the text, extracting KEs from academic papers, and using them as the main operands for the measurements have become the new frontier for measuring scholarly impact. According to the definition of Wang et al. (2019), a KE, such as a definition, theorem, rule, or method, is the smallest integral knowledge object in scientific literature. MKEs are usually considered as one of the most essential knowledge elements among all KEs (Hua, 2016; Chen & Kanuboddu, 2021). Its degree of innovation largely determines the quality of scientific papers (Wang et al., 2017; He & Du, 2020).Therefore, our research aims to develop a novel approach to detect and calculate the innovation degree of MKEs in scientific literature. To achieve this research goal, the following core issues need to be solved:

- **How to extract MKEs from academic papers?** The existing evaluation measures pay little attention to the innovation degree of specific knowledge content in scientific literature at the micro level of KEs. Therefore, it is impossible to know exactly where academic papers have made innovations. For this problem, we provide an effective approach that can extract MKEs from scientific literature automatically. Specifically, rule-based methods is used to extract MKEs from scientific literature.
- **How to represent MKEs?** The key challenge to derive innovation measure from the textual information of scientific literature is how to represent the semantics of MKEs effectively and efficiently without information loss. However, most of the existing quantitative evaluation measures treat the evaluation of scientific literature as a definite mathematical model, ignoring the fuzziness and randomness in the process of scientific literature assessment. Some evaluation models based on fuzzy set theory have indeed been proposed to solve the problem, but they lack comprehensive considerations of the randomness and fuzziness inherent in assessing the quality of scientific literature Behret and Gumussoy (2012). To this end, we establish a novel multi-dimensional cloud model to comprehensively characterize the randomness and fuzziness of MKEs. Based on this cloud model, we quantified the semantic changes of MKEs and used it as a proxy to measure the innovation in scientific literature.
- **How to calculate the innovation degree of MKEs?** Previously, scholars have used various methods for measuring innovation in scientific literature. However, until now, there is no academic consensus on the definition of innovation. Some scholars have discussed innovation as being related to impact Costanzo and Sánchez (2019) while others relate it to novelty (Packalen & Bhattacharya, 2019; Trapido, 2015). Since impact assessment is just a partial proxy for scientific literature's innovation and is mainly used at the macro level, this paper takes novelty as a measure of innovation in scientific literature. Specifically, we rely on the similarity between MKEs to measure the innovation of MKEs. Under the same research field, the higher the similarity between a MKE and other MKEs, the lower the innovation degree.

The remainder of the paper is organized as follows: "Related works" Section reviews the related work on measuring the innovation in scientific literature. "A framework for measuring the innovation of MKEs" Section describes the framework to measure the innovation degree of MKEs in scientific literature, including the extraction of MKEs, the generation of MKC drops, the generation of the MKC, and the calculation of the innovation degree of MKEs based on a similarity cloud. The experiments and results are presented in " Experiments and results" Section. Finally, we conclude the paper and discuss the future work in "Conclusion and future work" Section.

Related works

The evaluation of scientific literature and the measurement of the innovation degree of the scientific literature have been discussed in existing studies. The approaches can be summarized into two categories: metadata-based bibliometric measures and content-based measures.

Metadata-based bibliometric measures

The most popular measure is the bibliographic analysis based on the metadata record of scientific literature, including information about journals, citations, keywords, authors, and their affiliations.

The journal impact factor (JIF) (Garfield, 1999) is the earliest measurement Garfield (2006) and is now viewed as a well-established indicator of the scientific quality of scientific literature. The JIF method assumes that the articles published in high impact factor journals should be high quality Sombatsompop et al. (2006). One of the drawbacks of the JIF method is that a journal's high impact factor may not result from the citations of all articles; instead, it may merely be attributed to a small number of highly cited articles (Campbell, 2008; Colquhoun, 2003; Garfield, 2001). Therefore, we cannot equate the quality of scientific literature with the journal's impact factor in which it was published (Notkins, 2008; Uzzi et al., 2013).

To solve the problem of the JIF method, Frank proposed using citation counts to measure the quality of scientific literature (Frank, 2003). Citation counts can quantify both the use and impact of the cited scientific literature. In the present literature, a number of measures have emerged in relation to citation counts to quantify the research impact of scientific literature. Such approaches for measuring innovation could be supported by the suggestion that many new ideas in science are inspired by previous studies Zeng et al. (2017). However, these approaches can also be criticized since citation counts are accumulated over time. For example, an academic paper that received 50 citations in ten years is not necessarily better than one that received 40 citations in five years. Therefore, any citation-based metrics without removing the time factors are unreliable for measuring the quality of scientific literature (Kosmulski, 2011). In addition, the above metrics are not comprehensive enough as they only consider the numbers of citations. Instead, it is more reasonable to use weighted citations to assess scientific literature (Cai et al., 2019; Wang et al., 2020a).

However, citation metrics do not always appropriate for the assessment of scientific literature since: (1) Citation metrics only measure one aspect of an academic paper (impact or utility) and they alone cannot capture the distinct types of research contribution (Wu et al., 2019). (2) The citation count of an article cannot be directly applied for innovation measurement (Zhai et al., 2018) because citation count is influenced by various factors not directly related to the innovation in scientific literature (Onodera & Yoshikane, 2015).

An improved measurement "h-index" has been proposed (Hirsch, 2005). The h-index is an index that attempts to measure productivity and impact of scientific literature of a scientist or scholar. Investigations show that the h-index works well in some contexts (Bornmann et al., (2008); Lovegrove & Johnson, 2008), but can be invalidated by bias because of self-citations. Also, in the h-index-based measurement, a newly published article cannot easily be evaluated on quality since it has only, so far, achieved a lower citation rate because of its more recent publication (Jin et al., 2007; Rousseau & Leuven, 2008).

In addition to journals, citations, and authors, keywords - representing the subject matter of articles—are also an important entity of metadata. Keywords-based approaches believe that high thematic novelty of academic papers is associated with high innovation. Uddin and Khan (2016) used the combination of usual and unusual keywords by authors to identify innovation in academic papers. A similar measure of "thematic novelty" based on the rareness of keyword combinations has also been proposed.

The above metadata-based bibliometric measures mainly evaluate the quality (the same method used to evaluate impact) of scientific literature based on external data (such as journals, citation counts, and keywords), rather than focus on the content of the paper. Another drawback of the metadata-based bibliometric measures is their less semantically interpret-ability. Instead, using content-based measures will be more reliable and solid.

Content-based measures

With the development of content analysis in academic literature, KEs, such as research methods, experiments, and result analyses, are considered as more appropriate resources for measuring innovation in academic literature. Therefore, content-based measures have attracted increasing attention recently.

Peer review is one of the most frequently used methods for innovation measurement based on a paper's content. In this method, one or multiple domain experts are invited to evaluate the quality of a paper based on its content with several given criteria (Darling, 2015). Authors are required to revise and improve the content based on reviewer feedback; the peer review process is a good assurance of the quality of academic literature.

However, there are two unavoidable issues in the peer-reviewed process: (1) with the explosion of submissions, finding an equivalent number of appropriate reviewers is becoming more difficult. In addition, reviewers may be assigned too many papers to review; too large a workload may also reduce the review quality. (2) Peer review can be quite subjective since opinions of different reviewers (sometimes reviewers may have markedly different backgrounds) on the same article can vary greatly; this may introduce the bias issue (Reinhart, 2009; Walker & Rocha da Silva, 2015).

Therefore, an automatic approach to identifying the contributions and innovations in scientific literature is beneficial and urgent. Efforts have been made by the research community to automatically extract semantic information (Marcondes & da Costa, 2016; Ronzano & Saggion, 2016; Tkaczyk et al., 2015), such as knowledge claims (Dahl, 2008; Myers, 1992), from the scientific literature. A knowledge claim is defined as a sentence summarizing the knowledge contribution that peers recognize in the field (Hunston, 1993). Although what constitutes knowledge claims is not entirely clear, which brings great challenges to identifying knowledge claims from scientific literature, these research studies also provide great inspiration for fine-grained evaluation of scientific literature. Based on these studies, our research is a step forward by proposing an approach to measure the innovation of MKEs in scientific literature at the micro level.

A framework for measuring the innovation of MKEs

Although we do not know exactly how many kinds of knowledge claims there are in scientific literature, the MKE is believed to be an essential knowledge claim of scientific literature. MKE is defined as following:

Definition 1 *(Method knowledge element):* MKE refers to actions to be taken to investigate a research problem and the rationale for the application of specific procedures or techniques used to identify, select, process, and analyze information applied to understanding the problem (Bryman, 2008; Chu & Ke, 2017; Wang et al., 2019). MEK can be represented as $< C, (S_1, S_2, S_i, \dots, S_n) >$, where C is the designation, representing an MKE, and S_i is the ith statement, describing the method definition, purposes, preconditions, functions, substeps, effects and others (Wang et al., 2019). For example, "A feature selection algorithm based on random forest (RFFS) is proposed. This algorithm adopts random forest algorithm as the basic tool, the classification accuracy as the criterion function." is a MKE extracted from Yao et al. (2014).

Due to the uncertainties (i.e., randomness and fuzziness) of these statements, MKEs are difficult to identify. There are two main challenges: (1) formally describing the qualitative concepts using natural language; and (2) transforming qualitative concepts to quantitative values. Several knowledge representation models have been proposed to tackle the two issues, yet they fail to consider the same model's randomness and fuzziness of knowledge. Recently, word embeddings such as Word2Vec Mikolov et al. (2013), Glove Pennington et al. (2014), TextCNN Chen (2015), and Bert Devlin et al. (2018), have been widely applied for learning the word representation of the concepts since they can capture the semantic and context information of the text. Word2Vec Mikolov et al. (2013), Glove Pennington et al. (2014), TextCNN Chen (2015) belong to the static word embedding, while Bert belongs to the contextual word embedding. Compared to static word embeddings, which only produce a fixed context-independent representation for each word, contextual word embedding such as BERT embedding produces the word-level representation based on the information of the entire sentence. Therefore, the same word could have a different representation if the word appears in different sentences. The Bert model has produced the best performance in many NLP tasks (Li et al., 2020). However, the word embedding-based models fail to generate high-quality vector representations for less frequent or new terms (Gupta et al., 2019), making them hardly reusable in computing the innovation degree of MKEs since We usually use new terms to describe new MKEs. The uncertainties, such as randomness and fuzziness of linguistic concepts also make the word embedding-based models challenge for MKEs identification. To bridge this gap, it is more feasible for a representation model which takes both the randomness and the fuzziness into consideration. However, most existing representative theories, for example probability theory, fuzzy sets theory, only deal with the fuzziness and randomness of linguistic concepts, respectively. Therefore, in this article, we combine the randomness and fuzziness for MKEs identification. Meanwhile, the cloud model (Li et al., 2009), an uncertain transformation model proposed to transform qualitative concepts to quantitative values, is applied to represent the MKEs. The cloud model represents uncertainty and fuzziness using membership degree or certainty. Cloud and cloud drops are the basic units, which are defined as following:

Definition 2 *(Cloud and cloud drops):* Assume that U is a quantitative numerical universe of discourse and C is a qualitative concept in U. If $x \in U$ is a random implementation of concept C, $\mu(x) \in [0, 1]$, standing for the certainty degree for which x belongs to C, is a random variable with stable tendency.

In the cloud model, the natural language concepts can be expressed by numerical features (Wang et al., 2011). Among these features, E_x is the expected value. E_n is the entropy, which is a general method to measure uncertainty. Typically, the larger the entropy is, the more difficult it is to describe the concept qualitatively and quantitatively. H_e is a measure of entropy, also called the entropy of entropy. Hyper-entropy reflects the cohesiveness of the uncertainty of all points representing qualitative concepts in the universe. Generally, the greater the super entropy, the greater the dispersion of cloud drops, the greater the randomness of membership and certainty, and the

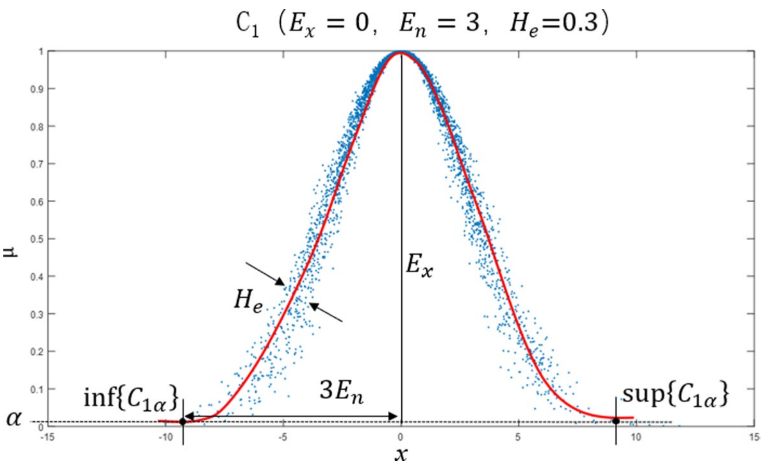

Fig. 1 An example of the normal form cloud

"thicker" the cloud thickness. Figure 1 shows the graph of a normal form cloud whose numerical characteristics are $E_x = 0$, $E_n = 3$, and $H_e = 0.3$.

The paper proposes a framework to measure the innovation degree of method knowledge elements (MIDMKE) based on the similarity cloud model. MIDMKE can be divided into four parts (which is shown in Fig. 2):

1. Extracting MKEs from the scientific literature.
2. Generating the cloud drops of each MKE based on BTM.
3. Generating the MKC of each MKE based on cloud drops.
4. Calculating the innovation degree of MKEs based on similarity cloud.

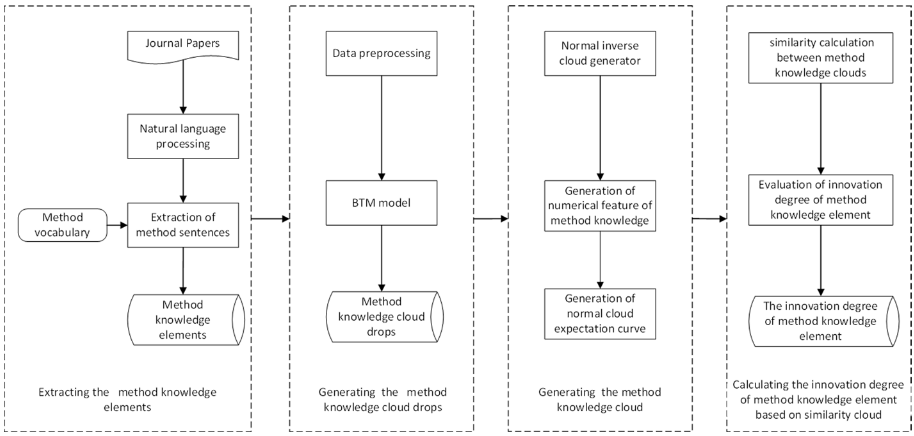

Fig. 2 The framework to measure the innovation degree of MKEs in a academic paper

Extracting the MKEs from scientific literature

MKEs consist of natural language sentences that are used to describe method knowledge. Therefore, the extraction of MKEs can be achieved by extracting method sentences from scientific literature. Rule-based methods are the most frequently used for sentence extraction. In our research, we reuse these methods to extract MKEs from scientific literature. Algorithm 1 presents the pseudo-code of the whole process and Fig. 3 depicts

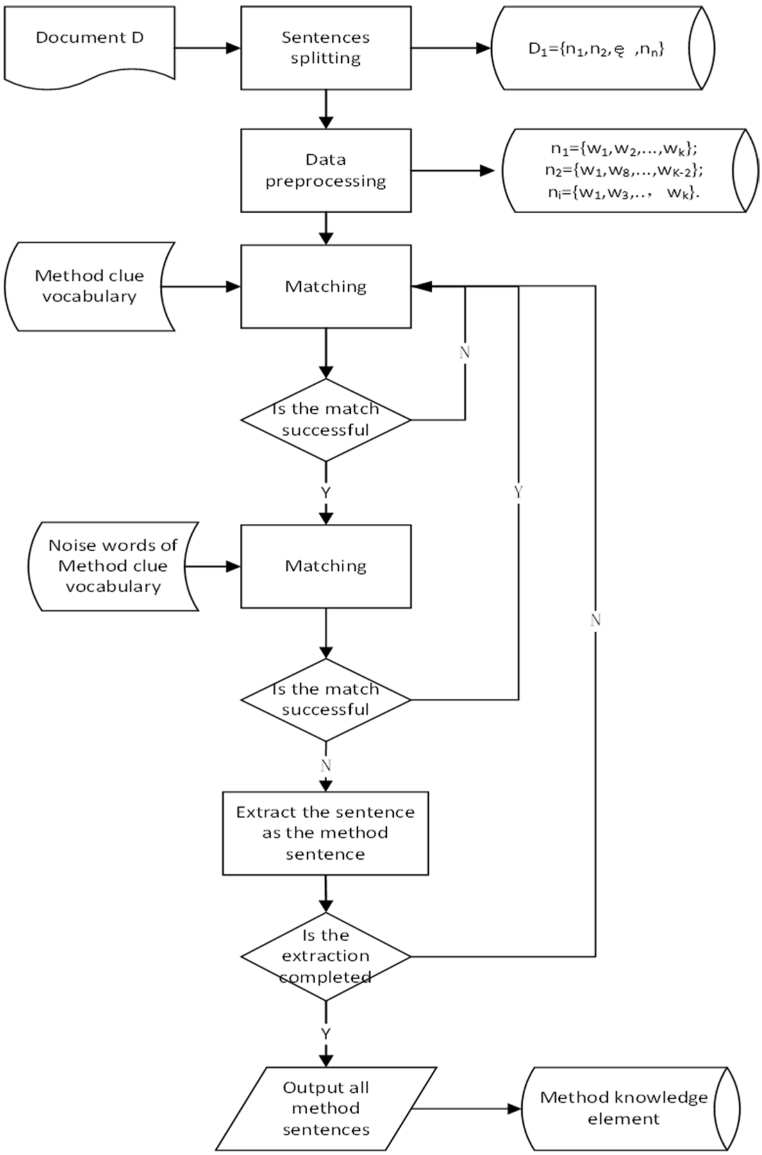

Fig. 3 The pipeline of method sentence extraction

the workflow of extracting MKEs. As for the rules for MKE extraction, We use two types of rules: method clue vocabulary and noise words of Method clue vocabulary, which are listed in Table 1.

Algorithm 1 Pseudo code for MKEs extraction

```
Input: An academic papers set as A = {D_1, D_2, ..., D_n}, n represents the number of academic papers of A.
for each i ∈ [1, n] do
    split academic paper D_i into sentences D_i = {S_1, S_2, ..., S_k} with the Stanford tokenizer, k represents the number of sentences in D_i.
    for each j ∈ [1, k] do
        preprocess sentence S_j includes doing word segmentation, removing punctuation marks and stop words, and getting a word vector S_j = {w_1, w_2, ..., w_p}, p represents the number of words in S_j.
        for each b ∈ [1, p] do
            search W_b in the method clue words collection, check whether the sentence S_k has method clue words or not.
            if search successfully then
                search W_b in the noise words of method clue vocabulary, check whether W_b is a noise word or not.
            else
                mark S_j as a sentence of MKE.
            end if
        end for
        extract S_j from the academic paper D_i.
    end for
    using all the extracted sentences to make up a MKE M_i.
end for
all the MKEs extracted from A_n.
Output: MKEs as B = {M_1, M_2, ..., M_m}, m represents the number of MKEs of B.
```

Generating the MKC drops

The method knowledge cloud is composed of each method knowledge cloud drop (MKCD) and all the MKCDs together indicate the characteristics of the MKEs. Thereby, in order to represent the MKEs using the cloud model, the first step is to generate the MKCDs (topic words that are used to describe the MKEs). For a MKCD x, its certainty degree $\mu(x)$ can be defined as the extent to which the drop can represent the MKE accurately. The more relevant a MKCD is to a topic, the greater the certainty degree of the MKCD.

Conventional topic models, such as PLSA and LDA, are among the most popular techniques for discovering topic words within a document (Zan et al., 2007; Lu et al., 2011). Since MKEs belong to short texts, applying these conventional topic models directly to such short texts usually does not achieve the expected performance. The reason is that conventional topic models implicitly capture the document level word co-occurrence patterns to reveal topic words and thus suffer from severe data sparsity in short documents. To tackle the sparsity problem of the conventional topic models, we use the BTM to identify topic words of a MKE in this paper. BTM is a word co-occurrence-based topic model that learns topics by modeling word-word co-occurrences patterns (Yan et al., 2013).

Table 1 The specific rules used to extract method KEs from the scientific literature

Type	Sub-type	Examples
Method clue vocabulary		
Vocabulary related	–	Method, means, way, technique, algorithm,
To method sentences		Step, procedure, measure, model, pattern, model set,
		Scheme, plan, program, project, etc.
Vocabulary related to method sentences	Define	⋯ be known ⋯
		⋯ the definition of ⋯ is ⋯
		⋯ be defined [as] ⋯
	Purpose	⋯ be usedl needed ⋯ to ⋯
		⋯ be to ⋯ determine identify validate provide ⋯
		⋯ to account analyses reduce ⋯
	Precondition	⋯ according since because ⋯
		⋯ be required to ⋯
		⋯ be based ⋯
		⋯ base on ⋯
	Function	⋯ using ⋯ result resulted in ⋯
		⋯ contribute lead to ⋯
		⋯ be help sue for ⋯
	Substeps	⋯ be ⋯ detected isolated calculated ⋯
		⋯ inferred ⋯ using
		⋯ be amplified aligned ⋯ by using ⋯
	Result	⋯ be found ⋯
		⋯ showed predicted ⋯
	Effect	⋯ confirm normalized ⋯
		⋯ shrinking the ⋯
		⋯ demonstrated demonstrate verified ⋯ that ⋯
		⋯ indicate reveal suggest that ⋯
Noise words of method clue vocabulary		
		rail high way, safe mode, civil procedure, sentence pattern, etc.

The initial idea of the BTM is to use the word pairs generated in the whole corpus to learn the topic of short texts and reduce the dimension of the document representation matrix by giving the number of specified topics and the number of words under each topic so that the words in the matrix can better represent the document (Tang, 2014). In this paper, the generation process of MKCDs based on the BTM is shown in Fig. 4.

There are three main steps:

1. Text preprocessing: Each MKE is tokenized during this step, and then stop words are removed.
2. BTM modeling: This step mainly includes word pair generation and parameter reasoning. The process of word pair generation is shown in Fig. 5, and the process of parameter reasoning is shown in algorithm 2. Two multinomial parameters, $\boldsymbol{\Phi}$ and θ, are obtained according to the counter and co-occurrence of word to topic assignment, where $\boldsymbol{\Phi}$ refers to the distribution of topic in the dataset of MKEs and θ refers to the global topic distribution in the dataset of MKEs. The specific calculation formulas are as follows:

$$\boldsymbol{\Phi}_{w|z} = \frac{n_{w|z} + \beta}{\sum_w n_{w|z} + M\beta}, \tag{1}$$

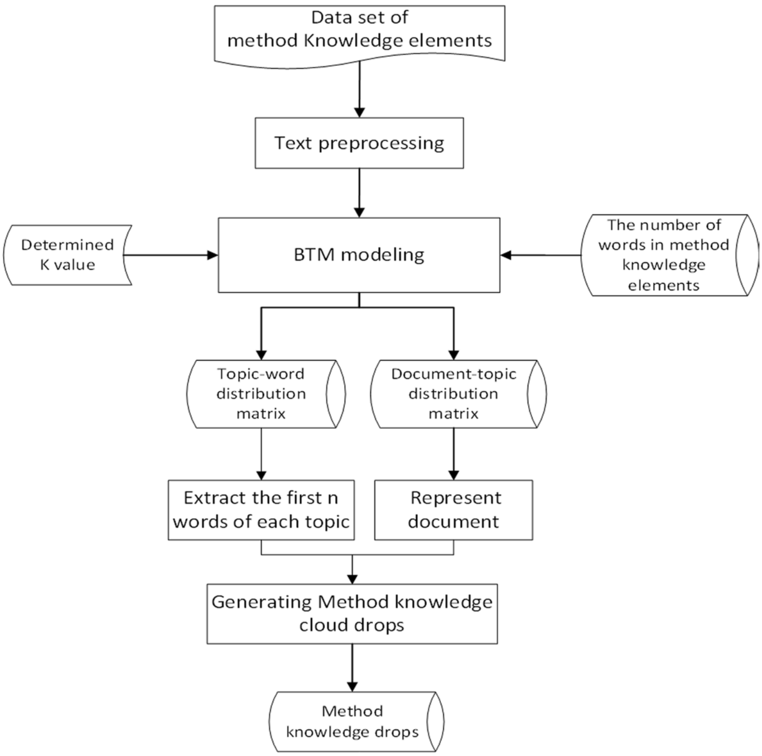

Fig. 4 The generation process of MKC based on BTM

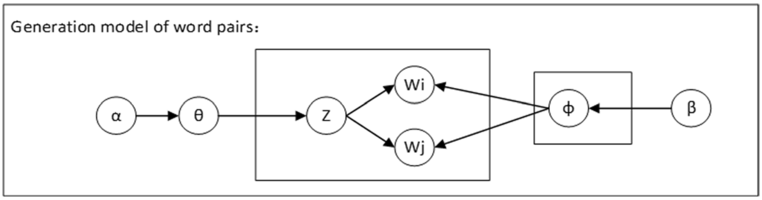

Fig. 5 The generation model of BTM word pairs

$$\theta_z = \frac{n_z + \alpha}{|B| + K\alpha}, \tag{2}$$

Where $\Phi_{w|z}$ refers to the probability of a word W in a topic z, θ_z refers to the probability of a topic z, $|B|$ refers to the total number of word pairs, α and β are the dirichlet priors, n_z is the number of times the biterm b is assigned to the topic z, and $n_{w|z}$ is the number of times the word w is assigned to the topic z. After the BTM modeling process, two output documents, the topic-word distribution matrix document and the document-topic distribution matrix document, are generated.

3. MKCD generation: the topic representation of each document (a MKE) is obtained using the document-topic distribution matrix, and then the first n words of each topic are extracted based on the topic-word distribution matrix as the MKCDs. Finally, the cloud drops of each MKE are obtained. Note that the cloud drops of i^{th} MKE is represented by $(w_{i1}, w_{i2}, \dots, w_{ij}, \dots, w_{in})$.

Algorithm 2 Reasoning algorithm of BTM parameters.

Input: topic number k, super parameters α and β, set of word pairs B.
initialization randomly arranging a topic for each word pair in the set of word pairs
for iteration = 1, 2, ..., n **do**
  for $b \in B$ **do**
    assign a topic Z_b to each word b pair according to $P(z|z_b, B, \alpha, \beta)$
    update $n_z, n_{w_i|z}, n_{w_j|z}$
  end for
  calculate Φ and θ according to formulas (1) and (2).
end for
Output: polynomial parameters Φ and θ.

Generating the MKC

Normal distribution is essential in probability theory, which is generally expressed by mean and variance. In fuzzy set theory, the bell membership function is widely used. The bell membership function is defined as:

$$\mu(x) = e^{-\frac{(x-a)^2}{2b^2}}, \tag{3}$$

A normal cloud is the most important model in cloud models (Hy et al., 2016). The normal cloud model is a representative model of conceptual uncertainty based on the normal distribution and bell membership function. The MKC is generated by two steps, which are described as follows:

1. Generating the digital characteristics of a normal cloud. Through the backward normal cloud generator (BNCG), given a limited set of MKCDs, the three digital characteristics *Ex*, *En*, and *He* could be produced to represent the corresponding MKE. The three cloud model digital characteristics of the *ith* MKE are calculated according to the following formulas.

$$Ex_i = \frac{1}{n}\sum_{j=1}^{n}\mu(w_{ij}), \tag{4}$$

$$En_i = \sqrt{\frac{\pi}{2}} \times \frac{1}{n}\sum_{j=1}^{n}|W_{ij} - Ex_i|, \tag{5}$$

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(W_{ij} - Ex_i)^2, \tag{6}$$

$$He_i = \sqrt{S_i^2 - En_i^2}, \tag{7}$$

Among them, n represents the number of cloud drops corresponding to each MKE; the sample variance S_i represents the degree of association between the core words and related words. The generation process of MKC based on the BNCG is shown in algorithm 3.

Algorithm 3 Generation process of MKC based on the BNCG.

1: **Input:** cloud drops of the i_{th} MKE, $(w_{i1}, \mu(w_{i1})), (w_{i2}, \mu(w_{i2})), \ldots, (w_{in}, \mu(w_{in}))$.
2: calculate the mathematical expectation Ex_i of cloud drops according to formula 4, and Ex_i is the most representative cloud drop in the MKC
3: calculate the entropy value En_i according to formula 5, which is a measure of the uncertainty of cloud drops, reflecting the degree of dispersion and the range of cloud drops
4: calculate the hyper-entropy value He_i according to formulas 6 and 7, which is the uncertainty measure of the entropy
5: **Output:** three digital characteristics (Ex_i, En_i, He_i) of the i^{th} MKE.

2. Generating the expectation curve of a normal cloud. Due to the obvious geometric characteristics of a normal cloud, we use the expectation curve to study the overall characteristics of MKC. The definition of a normal cloud expectation curve of MKC is as follows:

Definition 3 *(The normal cloud expectation curve of MKC):* Given the quantitative universe U, C is a MKE based on U. If the quantitative value in U is a random realization of C and satisfies the condition $x \sim (Ex, En^2)$, where $En \sim N(En, He^2)$ and $En \neq 0$, then the normal cloud expected curve of a MKE is as follows:

$$y = e^{-\frac{(x-Ex)^2}{2En^2}}, \tag{8}$$

In this paper, we use formula 8 to get the normal cloud expectation curve of each MKE.

Calculating the innovation degree of MKEs based on the similarity cloud

A new KE indicates the innovation of an article. Intuitively, a new KE can be identified by comparing it with KEs in the existing scientific literature: if it already exists in previous scientific literature, we cannot consider it a new KE; otherwise, it has a high possibility of being a new KE. Therefore, in this paper, we use the similarity between a MKE and other MKEs to measure the innovation degree of the MKE. Precisely, we first calculate the similarity between MKEs based on similarity cloud (Li et al., 2009; Wang et al., 2020b), then measure the innovation degree of a MKE through calculating the average similarity between it and other MKEs.

Calculating the similarity between MKEs

The similarity cloud algorithm (Li et al., 2009; Wang et al., 2020b) based on the expectation curve overlap of normal clouds is used to measure the similarity degree of MKCs. The overlap degree is proposed to describe the overlapping part of two clouds. The basic idea of the similarity cloud algorithm is to calculate the area that is the overlap between the expectation curve of MKCs $C_1(Ex_1, En_1, He_1)$ and $C_2(Ex_2, En_2, He_2)$ and judge the similarity degree between MKCs according to the size of the area. The specific calculation process is as follows:

1. Define the boundary of each MKC. There is a 3En rule that is 99.74% of cloud drops will fall on the interval $[Ex - 3En, Ex + 3En]$ in a normal cloud model. The cloud drops located outside of $[Ex - 3En, Ex + 3En]$ are called the small probability event. These cloud drops do not affect the overall characteristics of the cloud model if we do not consider them. Instead, we only need to consider the cloud drops distributed in this region when computing the similarity between MKCs. The boundary of MKC C_i is as follows:

$$\text{boundary}(C_i) = [Ex_i - 3En_i, Ex_i + 3En_i], \tag{9}$$

2. Define the overlap between method MKCs. There are two kinds of overlap between MKCs: one is that there is no overlap, the other is that there is overlap. There is no overlap between MKC C_1 and C_2 when the lower boundary of C_1 is greater than the upper boundary of C_2, or the upper boundary of C_1 is less than the lower boundary of C_2. In other words, the cloud similarity between C_1 and C_2 is 0. When there is overlap between C_1 and C_2, the formula of the overlap is as follows:

$$\text{overlap}(C_1, C_2) = [min(Ex_1 - 3En_1, Ex_2 - 3En_2), max(Ex_1 + 3En_1, Ex_2 + 3En_2)], \tag{10}$$

3. Calculate the intersection point between MKCs. Supposing there is an overlap between C_1 and C_2, the method of finding the intersection is as follows:

$$x_1 = \frac{Ex_2En_1 - Ex_1En_2}{En_1 - En_2}, y_1 = e^{-\frac{(x_1 - Ex)^2}{2En^2}}, \tag{11}$$

$$x_2 = \frac{Ex_1En_2 - Ex_2En_1}{En_1 + En_2}, y_2 = e^{-\frac{(x_2 - Ex)^2}{2En^2}}, \tag{12}$$

4. Calculate the similarity between MKCs. According to the relationship between the overlap and the intersection of MKCs C_1 and C_2, there are three kinds of similarity calculation methods:

   – The intersection points $(x_1, y_1), (x_2, y_2)$ are both outside the overlap of the MKCs C_1 and C_2. In other words, there is no overlap, so the cloud similarity between C_1 and C_2 is as follows:

   $$\mathrm{Sim}(C_1, C_2) = 0, \tag{13}$$

   – One of the intersections (x_1, y_1) and (x_2, y_2) is within the overlap of the MKCs C_1 and C_2. In other words, there is a cross relationship between MKCs C_1 and C_2. The cloud similarity between C_1 and C_2 is then as follows:

   $$\mathrm{Sim}(C_1, C_2) = \frac{\mu - \alpha}{1 - \alpha} \times \mathrm{overlap}(C_1, C_2), \tag{14}$$

   Where μ denotes the certainty degree of the intersection of MKCs C_1 and C_2, and α is the certainty degree of cloud model $3En$ rules.

   – The intersection points (x_1, y_1) and (x_2, y_2) are both within the overlap of the MKCs C_1 and C_2. In other words, there is an inclusion relation between MKCs C_1 and C_2 (see Fig. 6). The cloud similarity between C_1 and C_2 is as follows:

   $$\mathrm{Sim}(C_1, C_2) = \frac{\mu(max) - \alpha}{1 - \alpha} \times \mathrm{overlap}(C_1, C_2), \tag{15}$$

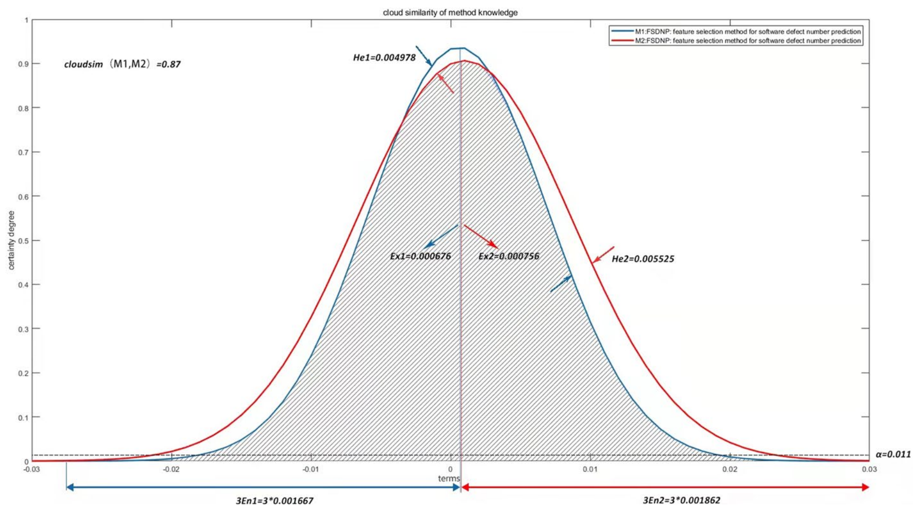

Fig. 6 Calculating the similarity between MKCs

where $\mu(max)$ denotes the certainty degree of the upper intersection point of (x_1, y_1) and (x_2, y_2).

Measure the innovation degree of MKE

In this paper, we rely on the similarity between MKEs to measure the innovation degree of the MKE. Under the same research field, the higher the similarity between the MKE C_i and other MKEs, the lower the innovation degree. Therefore, for a KE C_i, the formula to measure its innovation degree is as follows:

$$Cr(C_i) = \frac{\sum_{j=1}^{m}(1 - Sim(C_i, C_j))}{m}. \tag{16}$$

Experiments and results

Data

To validate the MIDMKE proposed in this paper, we conducted experiments on scientific literature of the “feature selection method (FSM)” research field in China. We selected the “FSM” as the test domain mainly for two reasons. First, the “FSM” is a research field mainly focusing on methods to reduce the number of input variables when developing a predictive model, academic papers of this research field are rich in MKEs. Second, the “FSM” is a subfield of machine learning. The total number of related papers is not very huge, so it is easier for domain experts to evaluate, interpret, and validate the results of our experiments.

To ensure that all possible relevant papers are collected, the top information provider in China, CNKI Data, is used. CNKI, whose purpose is a knowledge sharing throughout China and the world, is a key national project of China. Its China Academic Journals Full-text Database is the largest searchable full-text interdisciplinary Chinese journals database in the world. Finally, a total of 1191 academic papers of the “FSM” research field published in core journals in China were retrieved from CNKI Data. In our experiments, we split the 1191 academic papers into two sets: a reference set and a test set to evaluate how the MIDMKE proposed in this paper is working. The reference set includes 1072 academic papers used as the reference of academic papers, and the test set includes 119 academic papers recently published in different journals.

Baselines

– We first compare our rule-based MKE extraction strategy with TextCNN, a convolutional neural network (CNN) for text. In TextCNN, document matrices generated by word embeddings (ERNIE_Chinese, 12-layer, 768-hidden, 12-heads, 110M parameters) are input into a CNN to perform identification. In the convolutional layer, every filter performs convolution on each word and different feature maps are firstly generated. A max-overtime pooling operation is then used to select the most critical features, and the activation function is used in this step. Finally, the extracted features are concatenated as the penultimate layer and passed to a fully connected SoftMax layer to predict the probability distribution over class labels To avoid over-fitting, we use dropout and batch normalization on the penultimate layer and weight vectors to reduce the parameters, as recommended by Gong and Ji (2018).

– We also compare our rule-based MKE extraction strategy with Bert, the state-of-the-art (SOTA) for multiple NLP tasks (Devlin et al., 2018). Given a text together with its label as an input sequence, it will be converted using the pre-trained BERT model (Bert_base_chinese, 12-layer, 768-hidden, 12-heads, 110M parameters). We then fine-tune the model and add the simple softmax classifier to the top of BERT for MKE identification. The output is the probabilities of a phrase belong to the MKE.

Evaluation metrics

We use precision, recall, and F1-score as metrics to evaluate the performance on MKE extraction since they are the most used evaluation metrics for text classification (Li et al., 2020).

Experiment setup

MKE identification

We train the models on Windows11 21H2 machine with 1 NVIDIA GeForce RTX3070 Ti GPU(8G), 12th Gen Intel(R) Core(TM)(i7-12700k@3.61GHz) and 16GB of RAM. We set the batch size to 8, with a max sequence length of 128 and a learning rate of 2e-5 to ensure that the GPU memory is fully utilized. The dropout probability is always kept at 0.1. We use Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We empirically set the max number of the epoch to 15 and save the best model on the validation set for testing. We conduct five-fold cross-validation to avoid over-fitting. The epoch of TextCNN is set as 20.

Measuring the innovation degree of MKEs

We measured the innovation degree of MKEs in the academic papers of the test set. The process is as follows. First, MKEs were extracted from 1191 academic papers based on the propose rule-based algorithm. Second, MKCDs of each MKE were generated using the BTM (see Table 2), and then based on MKCDs, the MKC (*Ex*, *En*, and *He*) for each MKE is generated through the BNCG (see Table 2). Third, based on the digital characteristics *Ex* and *En* of each MKC, the corresponding normal expectation curve of the MKC is generated through formula 8 (see Fig. 7). Last, the similarity between MKCs is obtained by using the cloud similarity calculation formulas 13, 14, and 15. Then the innovation degree of each MKE is calculated by using the innovation degree calculation formula 16.

Results and analysis

Results on MKE identification

Table 3 presents the results of MKE identification of TextCNN, Bert, and the Rule-based method regarding precision, recall, and F-1 score. The rule-based method achieves the

Table 2 MKCs and their respective drops

MKE ID	MKCDs	Digital features of MKC
M1	Pattern recognition (0.0154), learning (0.1581),	(0.0043,0.0102,0.0204)
Zhang et al. (2005)	Classification (0.0590), clustering (0.1125), mean (0.0354),	
	Supervision (0.1786), impact (0.0347), correlation (0.0198)	
M2	Function (0.0304), space (0.0952), important (0.0562),	(0.0133,0.0286,0.0611)
Shang and Huang (2006)	Network (0.0121), performance (0.0945), automatic (0.0201),	
	Classification (0.5316), construction (0.0182)	
M3	Tradition (0.2077), category (0.2566), combination (0.1186),	(0.0153,0.0340,0.0496)
Yang et al. (2010)	Performance (0.0945), classification (0.5316),	
	Classifier (0.0932), Vector (0.0822), document (0.2319)	
M4	Distribution (0.1685), tradition (0.2077), category (0.2566),	(0.0138,0.0303,0.0490)
Pei and Liu (2011)	Classification (0.5316), precision (0.0528),	
	Corpus (0.0407), impact (0.0744), feature word (0.1751)	
M5	Information (0.2741), distribution (0.1685),	(0.0154,0.0351,0.0634)
Ren et al. (2010)	Probability (0.0122), tradition (0.2077), performance (0.0945),	
	Classification (0.5316), text (0.5442), gain (0.1772)	
M6	Criterion (0.0425), function (0.0271), subset (0.3759),	(0.0128,0.0293,0.0634)
Yao et al. (2014)	Random (0.1208), performance (0.1980),	
	Classification (0.6883), accuracy (0.1606), sequence (0.0199)	
M7	Information (0.1430), subset (0.3123), important (0.1117),	(0.0103,0.0229,0.0341)
Zhang et al. (2013)	Correlation (0.1344), performance (0.0911),	
	Classification (0.2090), classifier (0.0252), independent (0.0087)	
M8	Criterion (0.0623), matrix (0.0451), correlation (0.1344),	(0.0064,0.0143,0.0240)
Fan et al. (2013)	Tradition (0.0324), evaluation (0.1239), neural network (0.0026),	
	Classification (0.2090), variable (0.0142)	
M9	Search (0.0382), criterion (0.0230), inter-class (0.0241),	(0.0206,0.0406,0.0587)
Xie and Xie (2014)	Subset (0.1672), learning (0.0888), combination (0.0394),	
	Performance (0.1089), concentration (0.0208)	

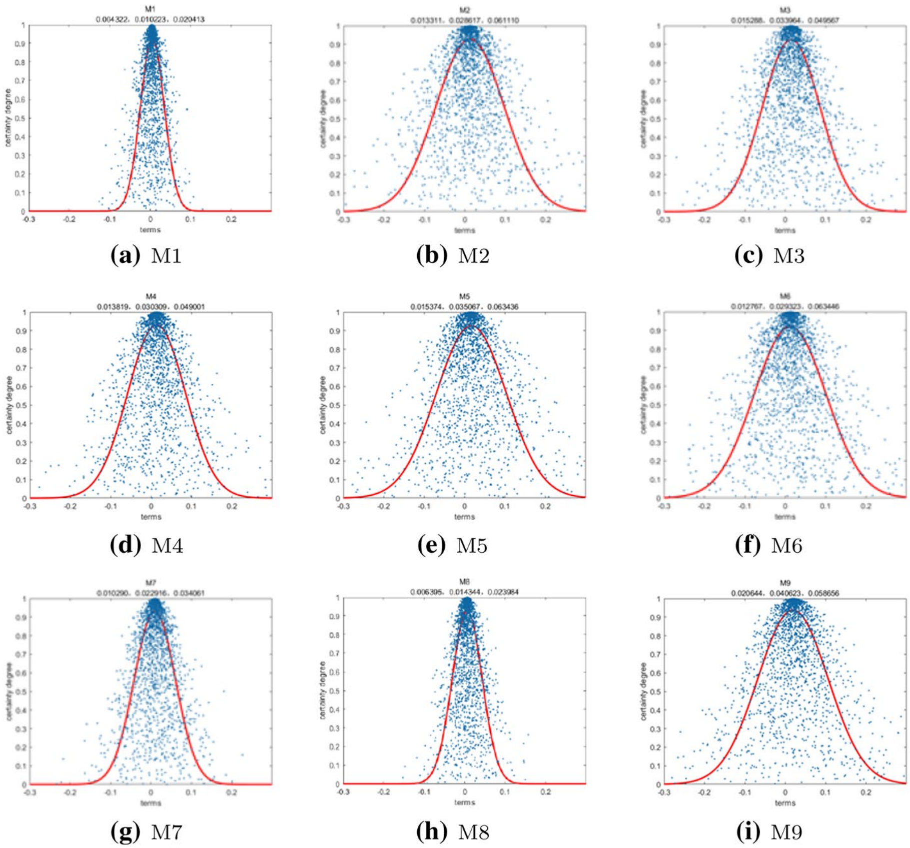

Fig. 7 **a** M1 **b** M2 **c** M3 **d** M4 **e** M5 **f** M6 **g** M7 **h** M8 **i** M9 Normal expectation curve of the MKC

Table 3 The performance of MKE identification

Model	Precision	Recall	F-1 score
TextCNN	0.9200	0.5610	0.6970
Bert	0.8000	0.9157	0.8539
Rule-based method	0.9527	0.9300	0.9412

best performance regarding all the three evaluation metrics, confirming the conclusion that deep learning models such as TextCNN and Bert need more data for training. When the dataset is small, traditional models such as the the proposed rule-based method can achieve a better performance.

Results on the innovation measurement of MKEs

The first issue for determining the performance of the MIDMKE is how the reference evaluation results (RER) are generated. Peer review, as a typical qualitative content-based analysis

Table 4 Evaluation results the innovation of MKEs

Rank	Article title	Journal	RER	Citation	MIDMKE
1	Unsupervised feature selection method based on K-means clustering	Application Research of Computers	9.5	88	0.9810
2	Research on feature selection in text classification algorithm based on Gini index	Application Research of Computers	9.8	121	0.9620
3	Feature selection method based on document frequency	Computer Engineering	8.6	150	0.8510
4	Research on improved CHI feature selection method in text classification	Computer Engineering and Applications	8.4	118	0.8280
5	Text feature selection method based on information gain	Computer Science	7.5	111	0.7750
6	Feature selection algorithm based on random forest	Journal of Jilin University (Engineering and Technology Edition)	8.5	450	0.7530
7	A multi-label feature selection algorithm based on information entropy	Journal of Computer Research and Development	7.0	130	0.7190
8	Feature selection algorithm for principal component analysis based on mutual information	Control and Decision	7.7	204	0.7280
9	Feature selection algorithm based on feature subset discrimination and support vector machine	Chinese Journal of Computers	7.9	145	0.7680

method for evaluating the innovation degree of academic papers, has been widely used to determine the innovation degree of academic papers. Due to this, we use peer review to obtain the RER, which serves as a gold standard in our experiments. Specifically, five experts familiar with the “FSM” research field are selected to score the innovation degree of MKEs in the academic papers of the test set on a scale of zero to ten. Then the judgments of these five experts are averaged to construct the RER (see Table 4). The higher the average score, the higher the innovation degree of the MKE.

The second issue for determining the performance of MIDMKE is related to choosing which index to use as the baseline index. Through analysis, we find that in the existing innovation degree measurement indices, both journal impact factor and h-index are indirect ways to measure the innovation degree of academic papers. On the one hand, the JIF wrongly equates the importance of a paper with the impact factor of the journal in which it was published; on the other hand, the h-index is mainly used to measure the impact of a particular scientist than an academic paper. Therefore, based on the above analysis, in order to evaluate the added value of integrating a cloud model into the innovation degree evaluation framework of an academic paper at the micro-level of KE, the citation is selected as the baseline index, which mainly focuses on the overall innovation degree of an academic paper at the macro level of literature and treats the evaluation of academic papers as a definite mathematical model, ignoring the fuzziness and randomness in the process of academic paper assessment. The experimental results are shown in Table 4.

Table 4 summarizes the rankings of the RER, citation, and MIDMKE indices. There are no obvious correlation distribution characteristics between these indices, so Pearson single-tailed correlation tests are conducted. The test results are shown in Table 5. The results show that RER and MIDMKE have a significant positive correlation ($1 > 0.922 > 0.5$) with a p-value of less than 0.01. It indicates that academic papers with low RER scores also have low MIDMKE scores, while those with high RER scores usually also have high MIDMKE scores. In theory, like MIDMKE, the citation indices should also have the same order as RER. However, in practice, there were no remarkable correlations between citation and RER. Because, although the citation index and the RER index have a medium correlation ($0.49 > 0.423 > 0.30$), its p-value is bigger than 0.05 ($0.071 > 0.05$). Many other studies have also demonstrated that the correlation between citation rate and score of peer evaluation is moderate Abramo et al. (2011); Mryglod et al. (2013). Through comparison, we can see that the MIDMKE proposed in this paper achieved better results than the citation index.

Conclusion and future work

Nowadays, with the rapid growth of published literature, it is necessary to automatically measure the innovation degree of scientific literature. Although the methods of the innovation degree measurement of the scientific literature have been improved from many aspects, they still face some limitations that need to be solved. In order to overcome these limitations, this paper proposes an innovation degree measurement framework of MKE in academic papers.

Table 5 Correlation analysis results

Variables	Significance (one-tailed)	Pearson
MIDMK - RER	0.010	0.922
Citation - RER	0.071	0.423

The main contributions of this study are twofold. First, we analyze the limitations of metadata-based bibliometric measures and content-based analysis measures, which are used to evaluate the innovation degree of scientific literature. The key to solving these problems is to evaluate the innovation degree of scientific literature at the micro-level of KE. Second, this paper proposes an innovation degree measurement framework of MKE in scientific literature to overcome these limitations. In this framework, we first provide an effective approach that can extract MKEs from scientific literature automatically, and then we establish a novel multi-dimensional similarity cloud model to characterize the randomness and fuzziness of MKEs comprehensively. Then we apply this model to determine the innovation degree of MKEs in academic papers. The effectiveness and performance of this framework have been validated through a case study.

However, the main defect of our study is the difficulty in fully extracting MKEs. The main reason for this difficulty is that it is impossible to obtain all the rules used in the rule-based MKEs extraction. In the future, we will conduct further studies on better MKEs extraction methods to improve the MIDMKE proposed in this paper. We will also integrate other knowledge elements using the cloud model for the innovation measurement.

Acknowledgements The authors are grateful to all the anonymous reviewers for their precious comments and suggestions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by ZW, KW, JL, JH, and HC. The first draft of the manuscript was written by ZW and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by Humanities and Social Science Research Foundation of Ministry of Education of China (Grant Number 21YJA870003) and National Social Science Foundation of China (Grant Number 19ZDA345).

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: A comparison of peer review and bibliometrics rankings. *Scientometrics, 89*(3), 929–941.

Behret, H., & Gumussoy, C. A. (2012). A fuzzy integrated approach for the selection of academic papers to a special issue. *International Journal of Applied Management Science, 4*(4), 371–384.

Bornmann, L., Wallon, G., & Ledin, A. (2008). Is the h index related to (standard) bibliometric measures and to the assessments by peers? an investigation of the h index by using molecular life sciences data. *Research Evaluation, 17*(2), 149–156.

Bryman, A. (2008). Of methods and methodology. *Qualitative Research in Organizations and Management, 3*(2), 159–168.

Cai, L., Tian, J., Liu, J., Bai, X., Lee, I., Kong, X., & Xia, F. (2019). Scholarly impact assessment: A survey of citation weighting solutions. *Scientometrics, 118*(2), 453–478.

Campbell, P. (2008). Escape from the impact factor. *Ethics in science and environmental politics, 8*(1), 5–7.

Chen, H., & Kanuboddu, B.N. (2021). A fine-grained annotation scheme for research contribution in academic literature. In *Proceedings of the 18th International Conference on Scientometrics and Informetrics,* pp 241–248.

Chen, Y. (2015). Convolutional neural network for sentence classification. *Master's thesis, University of Waterloo*.

Chu, H., & Ke, Q. (2017). Research methods: What's in the name? *Library & Information Science Research, 39*(4), 284–294.

Colquhoun, D. (2003). Challenging the tyranny of impact factors. *Nature, 423*(6939), 479–479.

Costanzo, B. P., & Sánchez, L. E. (2019). Innovation in impact assessment theory and practice: How is it captured in the literature? *Environmental Impact Assessment Review, 79*(106), 289.

Dahl, T. (2008). Contributing to the academic conversation: A study of new knowledge claims in economics and linguistics. *Journal of Pragmatics, 40*(7), 1184–1201.
Darling, E. (2015). Use of double-blind peer review to increase author diversity. *Conservation Biology, 29*(1), 297–299.
Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805.
Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013). Entitymetrics: Measuring the impact of entities. *PLoS ONE, 8*(8), 71–416.
Fan, X., Feng, H., & Yuan, M. (2013). Feature selection algorithm for principal component analysis based on mutual information (in Chinese). *Control and Decision, 28*(06), 915–919.
Frank, M. (2003). Impact factors: Arbiter of excellence? *Journal of the Medical Library Association, 91*(1), 4–6.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.
Garfield, E. (1999). Journal impact factor: A brief review. *CMAJ, 161*(8), 979–980.
Garfield, E. (2001). Impact factors, and why they won't go away. *Nature, 411*(6837), 522–522.
Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*(1), 90–93.
Gong, L., & Ji, R. (2018). What does a textcnn learn? arXiv preprint arXiv:180106287.
Gupta, V., Giesselbach, S., Rüping, S., & Bauckhage, C. (2019). Improving word embeddings using kernel PCA. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pp 200–208.
He, J., & Chen, C. (2018). Predictive effects of novelty measured by temporal embeddings on the growth of scientific literature. *Frontiers in Research Metrics and Analytics, 3*, 9.
He, L., & Du, X. (2020). Research on intelligent evaluation for the content innovation of academic papers (in Chinese). *Library and Information Service, 64*(1), 93–104.
He, Q., Chen, B., Pei, J., Qiu, B., Mitra, P., & Giles, L. (2009). Detecting topic evolution in scientific literature: How can citations help? In *Proceedings of the 18th ACM conference on Information and knowledge management*, pp 957–966
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences, 102*(46), 16569–16572.
Hua, B. (2016). Types and description rules of knowledge elements about method in academic papers (in Chinese). *Journal of Library Science in China, 42*(1), 30–40.
Hunston, S. (1993). Professional conflict-disagreement in academic discourse. In *Text and technology, John Benjamins,* p 115.
Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The r-and ar-indices: Complementing the h-index. *Chinese Science Bulletin, 52*(6), 855–863.
Kosmulski, M. (2011). Successful papers: A new idea in evaluation of scientific output. *Journal of Informetrics, 5*(3), 481–485.
Li, D., Liu, C., & Gan, W. (2009). A new cognitive model: Cloud model. *International Journal of Intelligent Systems, 24*(3), 357–375.
Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P.S., & He, L. (2020). A survey on text classification: From shallow to deep learning. arXiv preprint arXiv:200800364.
Lovegrove, B. G., & Johnson, S. D. (2008). Assessment of research performance in biology: How well do peer review and bibliometry correlate? *Bioscience, 58*(2), 160–164.
Lu, Y., Mei, Q., & Zhai, C. (2011). Investigating task performance of probabilistic topic models: An empirical study of PLSA and IDA. *Information Retrieval, 14*(2), 178–203.
Marcondes, C. H., & da Costa, L. C. (2016). A model to represent and process scientific knowledge in biomedical articles with semantic web technologies. *KO Knowledge Organization, 43*(2), 86–101.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:13013781.
Mörchen, F., Dejori, M., Fradkin, D., Etienne, J., Wachmann, B., & Bundschus, M. (2008). Anticipating annotations and emerging trends in biomedical literature. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp 954–962.
Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2013). Absolute and specific measures of research group excellence. *Scientometrics, 95*(1), 115–127.
Myers, G. (1992). 'in this paper we report...': Speech acts and scientific facts. *Journal of Pragmatics, 17*(4), 295–313.
Notkins, A. L. (2008). Neutralizing the impact factor culture. *Science, 322*(5899), 191–191.
Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology, 66*(4), 739–764.

Packalen, M., & Bhattacharya, J. (2019). Age and the trying out of new ideas. *Journal of Human Capital, 13*(2), 341–373.

Pei, Y., & Liu, X. (2011). Research on improved chi feature selection method in text classification (in Chinese). *Computer Engineering and Applications, 47*(04), 128–130.

Pennington, J., Socher, R., & Manning, C.D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),* pp 1532–1543.

Reich, Y. (1995). Measuring the value of knowledge. *International Journal of Human-Computer Studies, 42*(1), 3–30.

Reinhart, M. (2009). Peer review of grant applications in biology and medicine. Reliability, fairness, and validity. *Scientometrics, 81*(3), 789–809.

Ren, Y., Yang, R., & Yin, M. (2010). Text feature selection method based on information gain (in Chinese). *Computer Science, 39*(11), 127–130.

Ronzano, F., & Saggion, H. (2016). Knowledge extraction and modeling from scientific publications. In *International workshop on semantic, analytics, visualization*, Springer, pp 11–25.

Rousseau, R., & Leuven, K. (2008). Reflections on recent developments of the h-index and h-type indices. *COLLNET Journal of Scientometrics and Information Management, 2*(1), 1–8.

Shang, W., Huang, H., & Liu, Y. (2006). Research on feature selection algorithm based on gini index in text classification (in Chinese). *Application Research of Computers, 10,* 1688–1694.

Sombatsompop, N., Kositchaiyong, A., Markpin, T., & Inrit, S. (2006). Scientific evaluations of citation quality of international research articles in the sci database: Thailand case study. *Scientometrics, 66*(3), 521–535.

Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics, 107*(3), 1195–1225.

Tang, Q, (2014). Short text clustering method based on btm (in Chinese). *PhD dissertation, Anhui University*.

Tkaczyk, D., Szostek, P., Fedoryszak, M., Dendek, P. J., & Bolikowski, Ł. (2015). Cermine: Automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition (IJDAR), 18*(4), 317–335.

Trapido, D. (2015). How novelty in knowledge earns recognition: The role of consistent identities. *Research Policy, 44*(8), 1488–1500.

Uddin, S., & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics, 10*(4), 1166–1177.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science, 342*(6157), 468–472.

Walker, R., & Rocha da Silva, P. (2015). Emerging trends in peer review-a survey. *Frontiers in Neuroscience, 9,* 169.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416–1436.

Wang, J., Lee, Y. N., & Walsh, J. P. (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. *Research Policy, 47*(6), 1070–1083.

Wang, M., Zhang, J., Jiao, S., Zhang, X., Zhu, N., & Chen, G. (2020). Important citation identification by exploiting the syntactic and contextual information of citations. *Scientometrics, 125*(3), 2109–2129.

Wang, R., Gao, X., Gao, Z., Li, S., Gao, J., Xu, J., & Deng, W. (2020). Comprehensive reliability evaluation of multistate complex electromechanical systems based on similarity of cloud models. *Quality and Reliability Engineering International, 36*(3), 1048–1073.

Wang, S., Zheng, Z., Sun, Q., Zou, H., & Yang, F. (2011). Cloud model for service selection. In: *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS), IEEE,* pp 666–671.

Wang, Z., Shen, X., Huang, R., & Huang, J. (2019). Extracting method knowledge elements from scientific literature: A rule-based approach. *Proceedings of the Association for Information Science and Technology, 56*(1), 805–807.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378–382.

Xie, J., & Xie, W. (2014). Feature selection algorithm based on feature subset discrimination and support vector machine (in Chinese). *Chinese Journal of Computers, 37*(08), 1704–1718.

Yan, E. (2014). Research dynamics: Measuring the continuity and popularity of research topics. *Journal of Informetrics, 8*(1), 98–110.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, pp 1445–1456.

Yang, K., Zhang, Y., & Li, Y. (2010). Feature selection method based on document frequency (in Chinese). *Computer Engineering, 36*(7), 33–35.

Yao, D., Yang, J., & Zhan, X. (2014). Feature selection algorithm based on random forest (in Chinese). *Journal of Jilin University (Engineering and Technology Edition), 44*(01), 137–141.

Zan, W., Tsim, Y., Yeung, W., Chan, K., & Liu, J. (2007). Probabilistic latent semantic analyses (plsa) in bibliometric analysis for technology forecasting. *Journal of Technology Management & Innovation, 2*(1), 11–24.

Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports, 714*, 1–73.

Zhai, Y., Ding, Y., & Wang, F. (2018). Measuring the diffusion of an innovation: A citation analysis. *Journal of the Association for Information Science and Technology, 69*(3), 368–379.

Hy, Zhang, Ji, P., Jq, Wang, & Chen, Xh. (2016). A neutrosophic normal cloud and its application in decision-making. *Cognitive Computation, 8*(4), 649–669.

Zhang, L., Sun, G., & Guo, J. (2005). Unsupervised feature selection method based on k-means clustering (in Chinese). *Application Research of Computers, 42,* 23–24.

Zhang, Z., Li, S., & Li, Z. (2013). A multi-label feature selection algorithm based on information entropy (in Chinese). *Journal of Computer Research and Development, 50*(06), 1177–1184.