Journal of Informetrics 16 (2022) 101282

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Combination of research questions and methods: A new measurement of scientific novelty

Zhuoran Luo[a,b], Wei Lu[a,b,*], Jiangen He[c], Yuqi Wang[a,b]

[a] School of Information Management, Wuhan University, Wuhan, Hubei, China
[b] Information Retrieval and Knowledge Mining Laboratory, Wuhan University, Wuhan, Hubei, China
[c] School of Information Sciences, The University of Tennessee, Knoxville, TN, USA

ARTICLE INFO

Keywords:
Scientific novelty
Novelty measurement
Combinational novelty
Deep learning

ABSTRACT

As critical building blocks of scientific research, research questions and research methods are put forward to reveal the nature of a publication's scientific novelty. Although existing studies have examined scientific novelty from multiple combination-based views, the temporal and semantic complexity of research questions and methods remains to be fully addressed. To remedy this, we introduce a new approach to measuring the novelty of papers from the perspective of question-method combination. Specifically, we demonstrated a life-index novelty measurement based on the frequency and age of question terms and method terms. Furthermore, by using deep learning and representation learning techniques, we proposed a semantic novelty measurement algorithm based on the semantic similarity of terms. By using the dataset of papers collected from ACM Digital Library for evaluation, the effectiveness of our methods was evaluated by case studies and statistical analysis. Our work innovatively integrates the age, frequency, and semantics of research methods and research questions that characterizes novelty in scientific publications.

1. Introduction

Scientific novelty is essential but hard to quantify. Science innovation and breakthroughs are the key driving force for scientific and technological advances (STA). As carriers of scientific and technological innovation, novel publications have the potential not only to become breakthroughs in themselves but also to foster subsequent breakthroughs that may have far-reaching effects (Criscuolo et al., 2017; Min et al., 2018). Though innovative publications tend to have a greater impact than conservative publications in the research community, they remain rare among the vast amount of scientific publications overall (Fortunato et al., 2018). This is due to uncertainty following the article's publication (Fleming, 2001; Wang et al., 2017)—it may be a blockbuster, or it may be ignored, reflecting the "high risk/high reward" nature of novel research (Arrow, 1972). In addition to its inherent qualities of uncertainty and high risk, novel research may face some resistance from existing scientific paradigms (Kuhn, 2012; Wang et al., 2017), as well as delayed recognition (Garfield, 1980; Van Raan, 2004) due to the longer time it takes to be discovered. These complex qualities of scientific novelty, along with differences in researchers' perceptions of innovation, greatly increase the difficulty of novelty quantification.

Despite these challenges, researchers have carried out explorations in measuring novelty in various ways. From the perspective of the origin of innovation, researchers have found that the most influential scientific discoveries are based primarily on combinations—especially atypical combinations—of traditional ideas from previous work (Tahamtan & Bornmann, 2018; Uzzi et al., 2013; Wang et al., 2017). Many researchers maintain that the main source of novelty is the new recombination of previously uncombined elements or the combination of established elements with new concepts (Fleming, 2001; Lee et al., 2015; Schumpeter & Backhaus, 2003;

* Corresponding author at: Wuhan University, China
E-mail addresses: zoraluo@whu.edu.cn (Z. Luo), weilu@whu.edu.cn (W. Lu), jiangen@utk.edu (J. He), yuqi.wang2014@whu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.joi.2022.101282
Received 1 March 2021; Received in revised form 11 February 2022; Accepted 17 March 2022
Available online 11 April 2022
1751-1577/

Zhang et al., 2019). That is, the balanced mixture of new and established elements is the safest means of producing scientific advances (Fortunato et al., 2018). From the perspective of network structure, researchers have investigated the common features of citation structures using a complex network approach (Amplayo et al., 2018; Kaufer & Geisler, 1989; Min et al., 2021; Shibata et al., 2007; Weis & Jacobson, 2021), aiming to explore the network structures of potentially high-impact and highly cited articles and thereby discover innovative scientific metrics latent in those structures. For a long time, the number of citations has been widely used to measure the novelty of a published paper (Min et al., 2018; Shibata et al., 2010). In general, the view of combination as a source of innovation has been widely applied. Whether the innovation is measured in terms of citing journals, co-citing papers, keyword co-occurrence, or network structure (another form of combining), these methods reflect a principle of combination.

However, there may be multiple motivations for researchers to cite a paper (Bornmann et al., 2008), and an article's novelty does not always correlate positively with the number of citations received (Chandonia & Brenner, 2006; Yan et al., 2020). In addition, journal or paper combination metrics do not consider the content of the article when measuring its novelty, which tends to produce biased evaluation results. Furthermore, novelty calculation methods based on keyword combinations (Yan et al., 2020) do not take into account the temporal factors of components and combinations, which are important for novelty. Meanwhile, to the best of our knowledge, there is still a lack of research on combinational novelty measurement from the perspective of semantics. The quantification of a paper's novelty in existing studies tends to use knowledge units such as keywords and subject words, failing to distinguish these from the perspective of semantic of those units.

Word representation learning methods based on word embedding have been proven effective in identifying the semantic commonalities and differences between different words (Wang et al., 2019), which is the key to the measurement of text novelty. With the advancement of fine-grained text mining and entity extraction techniques, studies targeting the semantic functions of scientific publications' terms provide new ideas for novelty measurement studies.

The current state of research, and the research gap noted above, urge us to think about how to use the key contents of a paper (e.g., research question terms and research method terms) from the perspective of combination, so as to fully measure the paper's novelty while considering the temporal, frequency, and semantic characteristics of words. To this end, we here explore a new perspective on paper novelty measurement based on a combination of research questions and methods. We propose two methods for calculating the novelty of scientific publications by taking their research question terms, research method terms, and the combinations thereof as objects of study, while considering their temporal, frequency, and semantic characteristics. The ACM (Association for Computing Machinery) database was chosen as the data source, with 204,224 papers from 1951 to 2018 in the database selected as our experimental data. From these, we obtained the question and method terms for all 204,224 articles using the existing model and counted the first occurrence of these terms, their combinations, and the accumulated frequency of occurrence. Then, 12,496 articles published in 2018 were selected as test data, and question novelty, method novelty, question-method combinational novelty, and article novelty were calculated using our two proposed methods. The test results were analyzed using visualization and case studies; furthermore, we tested the relevance and advantages of the two methods proposed in this paper using correlation tests and regression analysis. Finally, we analyzed the distribution of articles with different novelty types using the results of the novelty calculations.

In this paper, we made two main contributions.

1 First, we introduced and implemented new approaches to measuring a paper's novelty from the viewpoint of question-method combination. Specifically, we proposed a novelty measure based on the frequency and age of terms and a novelty measurement algorithm based on the semantic similarity of terms.
2 Second, we demonstrate the effectiveness of the two proposed methods using case studies and statistical analysis.

2. Literature review

This section introduces the foundations of novelty studies for question-method combinations. It is decomposed down into three main themes: novelty measurements of scientific publications, combinational novelty, and term functions identification. The first one describes the main methods of novelty measurement of scientific literatures, the second introduces the theory foundation of combinational novelty and related studies, and the third presents the definition of lexical function and studies in lexical semantic identification.

2.1. Novelty measurements of scientific publications

The measurement or detection of the novelty of scientific publications can be divided into two main categories: 1) methods that evaluate papers through sources of innovative adoption based on citation analysis; and 2) discovery of innovative content based on textual analysis.

To a certain extent, the novelty of a paper can be evaluated by its mutual citation relationships. The most typical method is to evaluate the novelty of a paper by the cited articles in the references or the interdisciplinary citation index of the cited journals (Hurd, 1992; Ponomarev et al., 2014; Tang, 2004). When the disciplinary properties of the literature are difficult to obtain, analysis using citation structure is deemed to shed light on innovation through the "absorption-output" perspective of the paper, thus becoming another solution for the calculation of paper novelty scores (Funk & Owen-Smith, 2017; Small, 2006; Soler, 2007). Akin to the idea of citation structure, the novelty evaluation of papers through citation networks draws on relevant theories of social complexity networks to reach similar evaluation conclusions (Chen et al., 2011; Ruibin, 2018; Shibata et al., 2010).

Novelty identification from the content of scientific publications can provide a new approach to novelty evaluation beyond citation analysis methods. Chen and Fang (2019) extracted n-gram candidates regarding academic papers in the medical field and evaluated novelty by comparing historical occurrences in databases. Besides calculating the proportion of new keywords, Uddin and Khan (2016), among others, added metrics such as the number and length of keywords, disciplinary diversity, and network centrality to create a comprehensive ranking of paper novelty. Packalen and Bhattacharya (2015) derived their novelty index by extracting a 2-gram or 3-gram sequence of literature topics and comparing the first appearance time of the topic with the publish time of the literature. On the other hand, Ding et al. (2001) use co-word analysis to achieve domain knowledge evolution detection by analyzing the phase transformation of domain vocabulary.

2.2. Combinational novelty

The creation of innovations is spurred when a prior innovation or its components are assembled into an original design (Jones, 2009; Mukherjee et al., 2016; Uzzi & Spiro, 2005). Schumpeter (1939), the pioneer of innovation theory, claimed that innovation is the “recombination of elements of production,” i.e. a “new combination” of production elements or production conditions. Nelson and Winter (1982) claimed that “the creation of any sort of novelty in art, science, or practical life—consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence.”

Previous studies have applied this combination-based view to the field of technology innovation and scientific novelty. Fleming (2001) measures how often and for how long a component or combination of components has been used previously. He uses this as a measure of component (combination) familiarity which drives recombination uncertainty in breakthrough patent creation. Strumsky and Lobo (2015) use a similar approach to Fleming to describe patents with combinatorial novelty, and they find that patents of “technological origin” are quite rare, accounting for less than 1% of patents. Combinatorial novelty, however, is quite common, accounting for about 30% of patents since 2005. Verhoeven et al. (2016) provide a comprehensive measurement of technological novelty across two dimensions: novelty in recombination and novelty in knowledge origins. For each dimension, the measures are proposed using patent classification and citation information. Uzzi et al. (2013) analyzed 17.9 million papers spanning all fields of science; their findings suggest that the most influential science is based primarily on conventional combinations of prior work, but with the intrusion of unusual combinations, and the latter type of papers are twice as likely to be highly cited works. (Wang et al., 2017) measure scientific novelty by examining whether a published paper combines reference journals for the first time. They find a potential bias in novelty from standard bibliometric measures, with novel studies showing greater variation in citations. In a departure from reference- or citation-based novelty measures, (Yan et al., 2020) use keywords of papers to measure new combinations and new components. They argue that both new combinations and new components have an inverted U-shaped effect on the number of citations to a paper.

2.3. Term functions identification

The lexical function, or term function, is a way of describing the semantic role of terms in a text. In an early study of such functions, Kondo et al. (2009) classified the content elements of article titles into four categories—HEAD, METHOD, GOAL, and OTHER—and constructed domain-oriented methodological/technical evolution paths. Nanba et al. (2010) then used support vector machines to identify and classify the terms “technology” and “impact” in patents and scientific literature. Gupta and Manning (2011) classified the lexical functions of academic literature into three types: topic, technology, and domain, which they then automatically identified; Huang and Wan (2013) suggested that the term functions could be classified into three classes: “method”, “task”, and “other”. Cheng et al. (2019) defined lexical functions of academic texts as the semantic roles assumed by terms in academic texts and classified academic vocabulary into two categories: 1) domain-independent lexical functions containing research questions and research methods, and 2) domain-related lexical functions containing cases, tools, metrics, and datasets. Based on the framework of lexical functions of academic texts, Lu et al. (2019) achieved the automatic identification of questions or methods for paper keywords by introducing deep learning techniques.

From the above-mentioned studies, on the one hand, there are few studies that measure innovation from the perspective of textual content, which is the direct content reflecting innovation of scientific publications. On the other hand, combinational novelty theory and related studies provide support for our paper, but there are few studies that consider both term combinations and their semantic functions. Therefore, in this work, we focus on how to measure the novelty of articles from the perspective of question-method combinations. The remainder of this paper proceeds as follows: section 3 presents the two proposed novelty calculation methods along with one article novelty type classification method. The empirical study and results analysis follow in section 4, and discussion of the results is given in section 5. Section 6 concludes the paper.

3. Methodology

3.1. Novelty measurement on life-index

Research questions and methods are considered necessary components in scientific publications (Heffernan & Teufel, 2018). Indeed, some articles may have more than one question word and method word. We focus on those core terms that best express the research connotation of the article. Specifically, we define the core problem term as the main question or topic studied in the research paper, and the core method as the main method or technique used to address the research question. Therefore, our research problem is simplified to a combination of a core problem term and a core method term. Compared with general terms, the life index of scientific terms has a more obvious trajectory of change, reflecting the developmental tendency and innovation trend of the discipline, to a certain extent. It has been found that the life cycle of terms based on emergence, growth, decline, and extinction can reflect the innovation diffusion stage of the literature (He & Chen, 2018).

In this paper, we demonstrate that combining term functions with their co-occurrence life-index may offer a new perspective for the analysis of the innovation stage of publications. Therefore, we propose a calculation method to measure terms' life-index based on the frequency and age of question terms and method terms. We introduce the concept of potential development year (PDY) proposed by Pimentel et al. (2014) and simplified by Lu et al. (2021), defined by the latter as the period from the first year of keyword publication to the current year. We adopt this simplified definition in this paper and define our life-index function as a natural logarithm of PDY. The life-index of the scientific term v in the literature document D is thus calculated as:

$$Lifeindex(v) = N(v) \times \ln(T_D - T_v + 1) \tag{1}$$

where a smaller index value represents a earlyer stage in the life-cycle period and a higher novelty score. T_D denotes the time when document D was published, T_v represents the time when v first appeared in the dataset, and $N(v)$ denotes the times that v appeared in the dataset during the period $[T_v, T_D]$.

Defining the question term as q and the method term as m, we calculate their respective life-index as follows:

$$Lifeindex(q) = N(q) \times \ln(T_D - T_q + 1) \tag{2}$$

$$Lifeindex(m) = N(m) \times \ln(T_D - T_m + 1) \tag{3}$$

The life-index of the question-method combination (q, m) of document D is then calculated as

$$Lifeindex(q, m) = N(q, m) \times \ln(T_D - T_{(q,m)} + 1) \tag{4}$$

A smaller index value indicates a younger age and higher novelty of the combination. $T_{(q,m)}$ denotes the first co-occurrence time of (q, m), and $N(q, m)$ represents the number of times (q, m) co-occurred in the dataset during the period $[T_{(q,m)}, T_D]$.

The larger the life-index of a term, the smaller its novelty. Therefore, we propose a novelty formula based on the life-index score:

$$LifeIndexNovelty = 1 - \frac{\ln(x_i + 1)}{\ln \max(x_i + 1)} (x_i > 0) \tag{5}$$

This provides a normalized question-method life-index novelty score from [0,1]; x_i denotes a given life-index from Eqs. (2–4), and $\max(x_i + 1)$ is the maximum value of $x_i + 1$, the range of i is [1,k], k demotes the number of papers.

3.2. Novelty measurement based on semantics

3.2.1. Word embedding model training

Novelty calculation methods based on term life-index combine the frequency and temporal properties of terms yet fail to capture the semantic differences between terms well. To compensate for this, we measured the novelty of the question, the method term, and their combinations from a semantic perspective. Word embeddings map words or phrases to vectors of real numbers (Mikolov et al., 2013); it has been shown to improve the performance of NLP (natural language processing) tasks when these vectors are used as the base input (Wang et al., 2019), and the similarity between vectors can reflect the semantic similarity between words. The representation of word embeddings include Word2vec (Mikolov et al., 2013), Glove (Pennington et al., 2014), Elmo (Peters et al., 2018), BERT (Devlin et al., 2018), etc. In 2018, Google proposed the BERT (Bidirectional Encoder Representations from Transformers) model using Transformer (Vaswani et al., 2017) to construct multilayer bidirectional encoding, and the word vectors trained by BERT are well suited to the task of text similarity computation. To measure the novelty of question-method combinations in academic papers at the semantic level, we propose a method that calculates the similarity of word vectors yield by BERT (see Fig. 1).

First, the full-text data of the articles included in the ACM database are selected as the training corpus, performing data pre-processing operations such as word segmentation, symbol filtering, stem extraction, then the corpus is trained using the BERT deep learning model to obtain the word embedding representation of each word. The cosine similarity of the word embedding representation of the words to be detected is then calculated to yield the similarity score. Finally, the similarity scores are used in our proposed novelty calculation formula to obtain the novelty scores of each question word and method word, as well as the novelty scores of the combinations. For the BERT-based word vector model training, we use the entire experiment corpus as the training data and set the number of BERT model steps to 300,000 and the training batch sample size to 16. The remaining parameters are set by default. We obtained the word vector representation model after training the model.

3.2.2. Semantic novelty score calculation

In this paper, we suggest the novelty of a single term is relative to all terms in the dataset, i.e., if the term has not appeared in the dataset before, it is a new term. We likewise consider a combination of terms to be new the first time the two words are combined. That is, the newness of a combination term is determined by the combination elements. A schematic of the combined semantic novelty calculation process is shown in Fig.

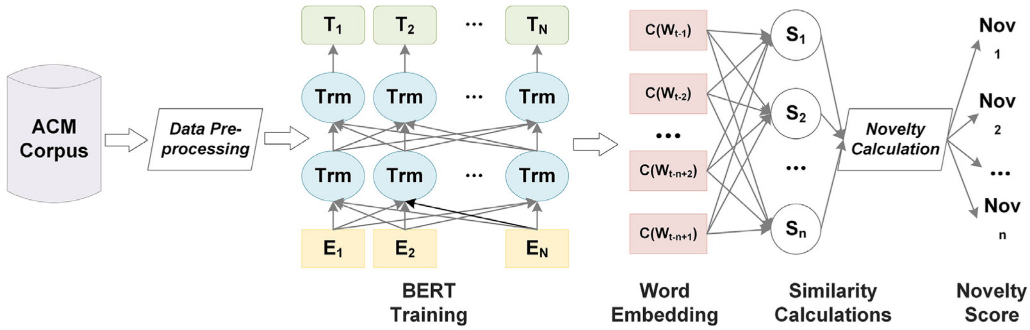

Fig. 1. BERT-based novelty calculation process for “question-method” combinations.

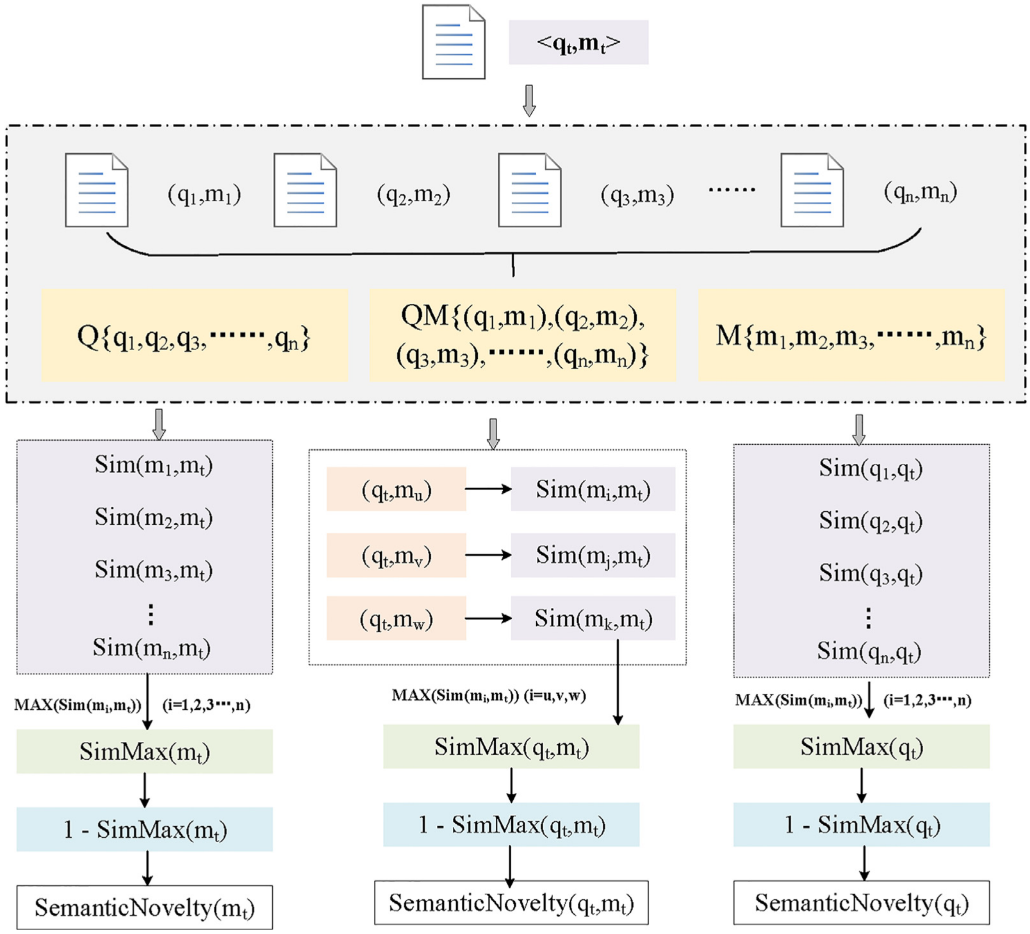

Fig. 2. Combinational semantic novelty calculation process. Q, M, and QM denote the sets of question terms, method terms, and question-method combinations in the historical dataset, respectively, and the elements in the sets are represented by their vectors; Sim is used to calculate the semantic similarity of two vectors; MAX denotes taking the maximum value.

As shown in Fig. 2, the novelty of the question terms and method terms are compared with all terms of that type in the historical dataset. That is, the vector similarity between the current question term q_t and each element of the historical question term set Q is calculated in turn, and the $SimMax(m_t)$ with the largest similarity score is subtracted from 1 to obtain the final question term novelty score. For the question-method combination, the novelty of the combination depends on the relative novelty of the combined terms. For example, an “old question + new method” combination is represented as $(q_t,\ m_t)$, where q_t is an old question term, and m_t is the new method term. In our algorithm, the current new method term m_t is relative not to all the method terms in our database, but to the combination method set, i.e., those method terms which have been combined with the target old question term q_t. That is, as long

Input: i, $article\_question(Q_i) = \{Q_1, Q_2, \ldots, Q_n\}$, $article\_method(M_i) = \{M_1, M_2, \ldots, M_n\}$, and ID of corresponding papers $article\_id(I_i) = \{I_1, I_2, \ldots, I_n\}$
Output: $semanticNoveltyScore(i)$

while i in $article\_id(I_i)$ do
  match $articleSet(I_i) \leftarrow delect\ i\ from\ article\_id(I_i)$;
  if $i\_question\ not\ in\ questionSet(Q_i)$ **then**
    $questionNovelty(i) \leftarrow 1$; // $i\_question$ is a new question
  else
    while $j\ in\ articleSet(I_i)$ **do**
      if $i\_question = j\_question$ **then**
        add $j\_method\ into\ methodSet(M_i)$;
        $V_m \leftarrow get\ i\_method\ Bert\ word\ vector$;
        $V_{M_i} \leftarrow get\ j\_method\ Bert\ word\ vector$;
        $S_{m_i} \leftarrow Sim(V_m, V_{M_i}) = \frac{V_m \cdot V_{M_i}}{\|V_m\| \|V_{M_i}\|}$;
        add S_{m_i} $into\ i\_methodBertSimilarity$;
      end
    end
    $questionNovelty(i) \leftarrow 1 - \max(i\_methodBertSimilarity)$;
    $semanticNoveltyScore(i) \leftarrow questionNovelty(i)$;
  end
  if $i\_method\ not\ in\ methodSet(M_i)$ **then**
    $methodNovelty(i) \leftarrow 1$; // i_method is a new method
  else
    while $j\ in\ articleSet(I_i)$ **do**
      if $i\_method = j\_method$ **then**
        add $j\_question\ into\ questionSet(Q_i)$;
        $V_q \leftarrow get\ i\_\ question\ Bert\ word\ vector$;
        $V_{Q_i} \leftarrow get\ j\_\ question\ Bert\ word\ vector$;
        $S_{q_i} \leftarrow Sim\left(V_q, V_{Q_i}\right) = \frac{V_q \cdot V_{Q_i}}{\|V_q\| \|V_{Q_i}\|}$;
        add S_{q_i} $into\ i\_questionBertSimilarity$;
      end
    end
    $methodNovelty(i) \leftarrow 1 - \max(i\_questionBertSimilarity)$;
    $semanticNoveltyScore(i) \leftarrow methodNovelty(i)$;
  end
  if $methodNovelty(i) = questionNovelty(i) = 1$ **then**
    $semanticNoveltyScore(i) \leftarrow 1$;
  end
  return $semanticNoveltyScore(i)$;
end

Algorithm 1. semanticNoveltyCalculate (Calculating the novelty of question-method combinations based on semantic similarity).

as the current method term does not appear in the historical method term set of the target question term, shown as $\{m_u, m_v, m_w\}$ in Fig. 1, the method is a new one for the combination. We then calculate the similarity between m_t and each element in $\{m_u, m_v, m_w\}$ and subtract the $SimMax(q_t, m_t)$ with the largest similarity score from 1 to get the novelty score of the question-method combination.

To calculate the semantic novelty, we first extracted the question terms and method terms of all articles in the historical dataset, and formed the sets of question terms, method terms, and question-method combinations, respectively. For a newly published paper, first, the target core question term and method term were extracted. These were then searched for in the historical database, and if either term was new, it indicated that it had not appeared in the dataset and had not been combined with other words. Combinations of new problem + new method terms have the highest novelty score (i.e., 1). For combinations containing old terms, we calculate novelty case by case. A more detailed description of the calculation is given in Algorithm 1.

Furthermore, since the similarity results of vector calculations vary widely, we normalized the calculated novelty scores from Algorithm 1 (without changing the magnitude of the scores) for better data presentation and analysis:

$$SematicNovelty = \frac{semanticNoveltyScore - semanticNoveltyScore_{\min}}{semanticNoveltyScore_{\max} - semanticNoveltyScore_{\min} + t} \tag{6}$$

Here, $SematicNovelty$ denotes the normalized semantic novelty score of the question term, method term, and question-method combination, with values in the range of (0,1); $semanticNoveltyScore_{\min}$ and $semanticNoveltyScore_{\max}$ denote the minimum and maximum values of the $semanticNoveltyScore$, respectively. To avoid a zero denominator in Eq. (6), a constant t is added to the denominator and its value is set to 0.0001.

3.3. Article novelty calculation and classification

After obtaining the novelty of the question terms, method terms, and question-method combinations of the paper, we further conduct two studies: i) calculate the overall article novelty, and ii) identify the article's innovation type.

The overall novelty of an article depends on the novelty of its components. When we look at the research question and research method as the main innovative elements, the $NoveltyScore$ of the paper can be calculated in two ways, one based on life-index and the other on semantic similarity:

$$LifeIndexNovelty(D) = (LifeIndexNovelty(q) + LifeIndexNovelty(m) + LifeIndexNovelty(q, m))/3 \tag{7}$$

$$SematicNovelty(D) = (SematicNovelty(q) + SematicNovelty(m) + SematicNovelty(q, m))/3 \tag{8}$$

When we investigate the question-method combinational novelty, both new combinations and new components are taken into account. If the novelty value of the article exceeds a certain threshold, the article can be considered novel, and the novelty types can be divided according to the novelty value. We set the novelty threshold T_{novel} to the median of the novelty scores for all terms, and the value depends on the distribution of novelty scores for all data. After judging the novelty of the question term and method term based on T_{novel}, the novelty of articles can be classified into four categories: new question + new method, old question + new method, new question + old method, and old question + old method. The latter can be further divided into two subcategories: new combination of old question + old method and old combination of old question + old method.

4. Empirical Analyses

4.1. Data preparation

As mentioned in 3.1, some articles may combine multiple research questions or methods. For our combinational novelty study, however, measuring the combination of core questions and core methods is enough to achieve our research goal, and the automatic extraction of multiple questions and methods is the next challenging task that remains to be solved. We introduced the question-method recognition model proposed by Lu et al. (2020), who trained the classification model BERT+LSTM with a training corpus of papers in computer science-related fields in 10 years to achieve the identification of research questions and research methods. The accuracy, recall, and F1 values of the experimental model reach 0.83, 0.87, and 0.85, respectively, and the results show that their proposed method is better than the traditional one.

Papers from 1951 to 2018 in the ACM database were selected as the experimental corpus. Then, data pre-processing methods such as word segmentation, stem extraction, punctuation filtering were used to clean the experimental corpus, and 204,224 papers were retained as our experimental data. Based on BERT+LSTM, we extracted question terms and method terms from the articles in the corpus. Since we only used the existing model to identify the question and method terms, the model would only output identification results. To examine the identification effect, we invited three graduate students to perform a manual test on 300 randomly selected articles, and the results showed that 241 papers (80.3%) were correctly recognized. The accuracy is slightly lower than that of the model, but it is an acceptable range because identifying different functions of words remains a more difficult task in natural language processing research. Next, the DOI, title, abstract, keywords, publication time, and other bibliographic information of each paper were extracted by using natural language processing methods. Finally, question terms, method terms and original literature fields are joined together and stored in the database.

We further counted the number of occurrences of each question-method combination and indexed each question term and method term in alphabetical order to build a question-method index table. Finally, 12,496 article records published in 2018 were extracted as data for analysis using the conditional query method in the database (seen as Fig. 3). The remaining 191,728 records were used as historical data to construct the question-method sample space.

4.2. Life-index novelty measurement

We first calculated the frequency and first occurrence time of the question terms, method terms, and their combinations in the test set of 12,496 articles. Then their life-index novelty was calculated according to formulas (1–5). We define the life-index novelty using the following abbreviations: LIN_Q for life-index novelty of question term, LIN_M for life-index novelty of method term, LIN_QM for life-index novelty of question-method combination, and LIN_D for life-index novelty of the document overall. Descriptive statistics of the life-index novelty scores of 12,496 are shown in Table 1, in which includes the maximum value, minimum value, mean and standard deviation of LIN_Q, LIN_M, LIN_QM and LIN_D. From the statistical data, it can be seen that the life index novelty interval of the question terms, method terms and their combinations is [0,1], and the combination LIN_QM has the largest mean value and the

pid	aid	title	question	method	publish year
1978942	1978947	Simulating the feel of brain-computer interfaces for design, developme	motor-imageri based	error-pron characterist	2011
2488388	2488525	Making the most of your triple store: query answering in OWL 2 using a	exact queri answer	off-the-shelf tripl store	2013
1228784	1228921	A fast clock scheduling for peak power reduction in LSI	peak power reduct	fast power estim method	2007
1877868	1877885	Human action recognition with MPEG-7 descriptors and architectures	human action recogni	mpeg-7 descriptor	2010
3122831	3122840	Locomotor: transparent migration of client-side database code	hybrid ship	static analysis	2017
2414536	2414599	Design considerations for after death: comparing the affordances of th	onlin enact of murder	onlin environ design	2012
1145706	1145711	LCARS: the next generation programming context	pervas applications	high-level graphic languag	2006
2736084	2736096	A logical memory model for scaling parallel multimedia workloads	fine-grain kernel sche	compil and runtim support	2015
1810617	1810671	Citation based plagiarism detection: a new approach to identify plagiar	plagiar detect	citat analysis	2010
3126594	3126660	Designing and Evaluating Livefonts	static learnabl of sigh	anim script	2017
3025453	3025745	Participatory Media: Creating Spaces for Storytelling in Neighbourhooc	neighbourhood plan	participatori media technolog	2017
2160125	2160131	A method to evaluate metal filing skill level with wearable hybrid senso	metal file	effect skill train	2012
1822090	1822172	Integrating categories of algorithm learning objective into algorithm vis	algorithm visual	learner-cent framework	2010
129712	129754	Symmetry and complexity	tight hierarchi	symmetri	1992
2909827	2930795	Practical and Scalable Sharing of Encrypted Data in Cloud Storage witl	sensor network	practic attribute-bas encrypt	2016
2984356	2985225	Running ModelGraft to Evaluate Internet-scale ICN	information-centr net	gamifi simulation	2016
2676723	2691943	"Maker Innovators": A Workshop for Youth Creating Responsive and W	comput and game	bidirect tangibl interfac	2015
1143997	1144292	Evolving musical performance profiles using genetic algorithms with st	express music perfor	genet algorithm	2006
3025453	3025497	The Effects of Artificial Landmarks on Learning and Performance in Sp	improv spatially-st gr	artifici landmark	2017
1364901	1364919	Direct sampling on surfaces for high quality remeshing	2d fast poisson samp	geodes distanc	2008
1143549	1143749	A wireless test bed for mobile 802.11 and beyond	wireless test bed	the feder railroad administr	2006
2016911	2016937	Learning web development: challenges at an earlier stage of computin	comput educ	content analysis	2011
74382	74487	Capturing designer expertise the CGEN system	computer-aid design	autom knowledge-acquisit to	1989
313237	313241	Survivable load sharing protocals: a simulation study	surviv wireless acces	dynam load balanc	1999
3106195	3106225	Detecting Variability in MATLAB/Simulink Models: An Industry-Inspired	matlab/simulink mode	customiz comparison proced	2017
3159450	3162345	Summit Selection: Designing a Feature Selection Technique to Support	noisi mix data	featur select	2018
3059454	3078699	Towards Lifelong Interactive Learning For Open-ended Embodied Narr	plan modeling collabc	lifelong interact learn	2017
330560	331013	Virtual synchronization: uncoupling synchronization annotations from s	thin film	sens and control	1998
2872427	2883010	Non-Linear Mining of Competing Local Activities	competit strong	multipl keyword	2016
3173574	3173854	Understanding Face and Eye Visibility in Front-Facing Cameras of Sma	commod mobil applic	face and eye detect	2018
2598153	2600049	Static hand poses for gestural interaction: a study	gestur interact	static hand pose	2014
3010915	3010989	Design and evaluation of a dynamic-interactive art system: a mixed me	an interact art systen	mix method	2016
3269206	3271726	CRPP: Competing Recurrent Point Process for Modeling Visibility Dyna	recurr point process	recurr neural network	2018
1460563	1460642	A wiki instance in the enterprise: opportunities, concerns and reality	wiki-bas application	researchwiki	2008

Fig. 3. Representative screenshot of experimental data.

Table 1
Descriptive statistics of life-index novelty scores.

	Minimum	Maximum	Mean	Std. Deviation
LIN_Q	0	1	0.8850	0.2046
LIN_M	0	1	0.8509	0.2401
LIN_QM	0	1	0.9977	0.0393
LIN_D	0.1709	1	0.9112	0.1056

smallest standard deviation, indicating that the novelty value of the question-method combination is large and concentrated. Based on the calculated scores, we produced a scatter plot of life-index novelty for the test data, shown in Fig. 4. Each point in the graph represents an article.

Fig. 4 shows that the overall distribution of points is converging toward the high-novelty positions, while the points in the low-novelty interval are more sparsely distributed, indicating that there are more articles with high novelty. The distribution is equal in terms of dot size, indicating that new combinations exist between old and new terms, and not just among new ones. In addition, it can be seen that the points on the upper and right edges of the graph are densely distributed and almost connected in a straight line, and the points are denser in the area near the upper left corner. This suggests that both new question terms and new method terms yield combinations with previous terms, while preferring to combine with new terms.

For a more intuitive understanding, we take four representative dots as examples for analysis; these are numbered and marked with red circles in Fig. 3. Table 2 shows each article's title, along with time, frequency, and novelty-score data for its question and method terms.

The first sample article proposes a novel type of ASE (artificial subtle expressions) that uses vibration (vibrational ASEs); both the question and the method are novel, with life-index novelty of 1. The second sample article investigates the analysis of the average load definition of "wireless sensor networks." The question term first appeared in 1970 and has accumulated 667 occurrences by 2018, with a novelty of 0. The method term is a new one in 2018, with a novelty of 1. It is a new combination of "old question + new method," and the life-index novelty of the paper is 0.667. The third article proposes different visualization techniques to solve the multi-objective optimization problem, the question term and method term are old, and combinations of the two have already appeared, with a lower life-index novelty of 0.615. The fourth article studies the use of genetic programming to enhance data to test the performance of feature selection more scientifically. The frequency of the question term is 53, that of the method is 216, and the combined novelty

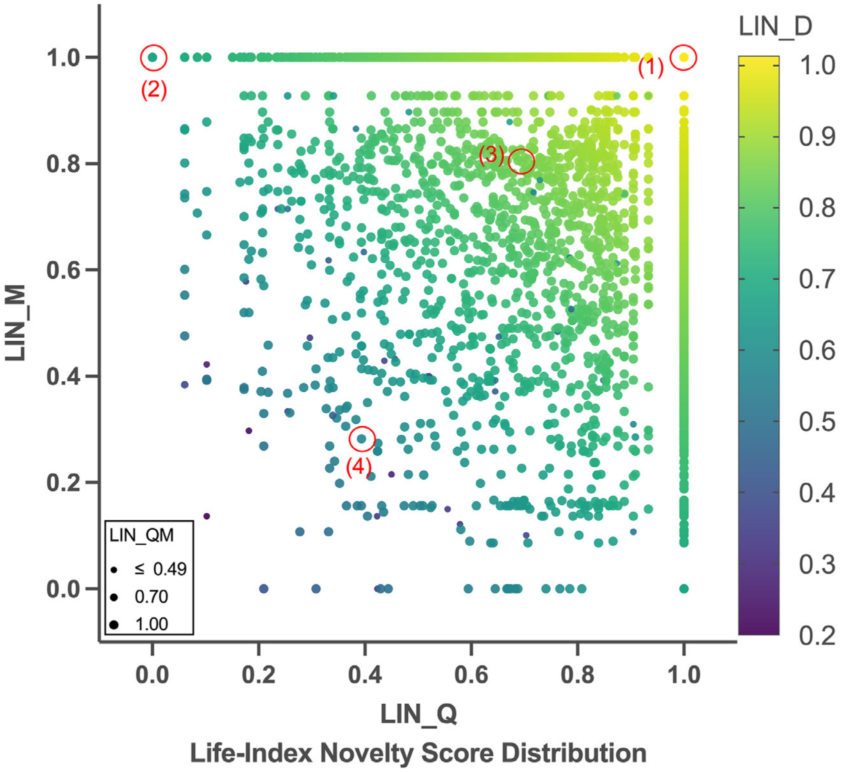

Fig. 4. Scatter plot of test papers' life-index novelty scores. x-axis represents LIN_Q, y-axis represents LIN_M, dot size (size of smallest symbol is set of 3) represents LIN_QM, and dot color indicates LIN_D.

Table 2
Life-index novelty results for 4 sample articles.

id	article title	question/ first-year/ times	method/ first-year/ times	LIN_Q/LIN_M/ LIN_QM/LIN_D
1	Vibrational artificial subtle expressions: Conveying system's confidence level to users by means of smartphone vibration	artificial subtle expression/ 2018/1	smartphone vibration /2018/1	1/1/ 1/1
2	Analysis of concise "average load" definitions in uniformly random deployed wireless sensor networks	wireless sensor network/ 1970/667	average load definition /2018/1	0/1/ 1/0.667
3	Visualising the search process for multi-objective optimisation	multi-objective optimization/ 2007/6	visualising technique /2007/4	0.434/0.575/ 0.836/0.615
4	Automatically evolving difficult benchmark feature selection datasets with genetic programming	feature select / 1997/53	genetic program /1984/216	0.393/0.121/ 0.491/0.335

Table 3
Descriptive statistics of semantic novelty scores.

	Minimum	Maximum	Mean	Std. Deviation
SN_Q	0	1	0.8828	0.1967
SN_M	0	1	0.8636	0.2046
SN_QM	0	1	0.7773	0.2276
SN_D	0	1	0.8412	0.1705

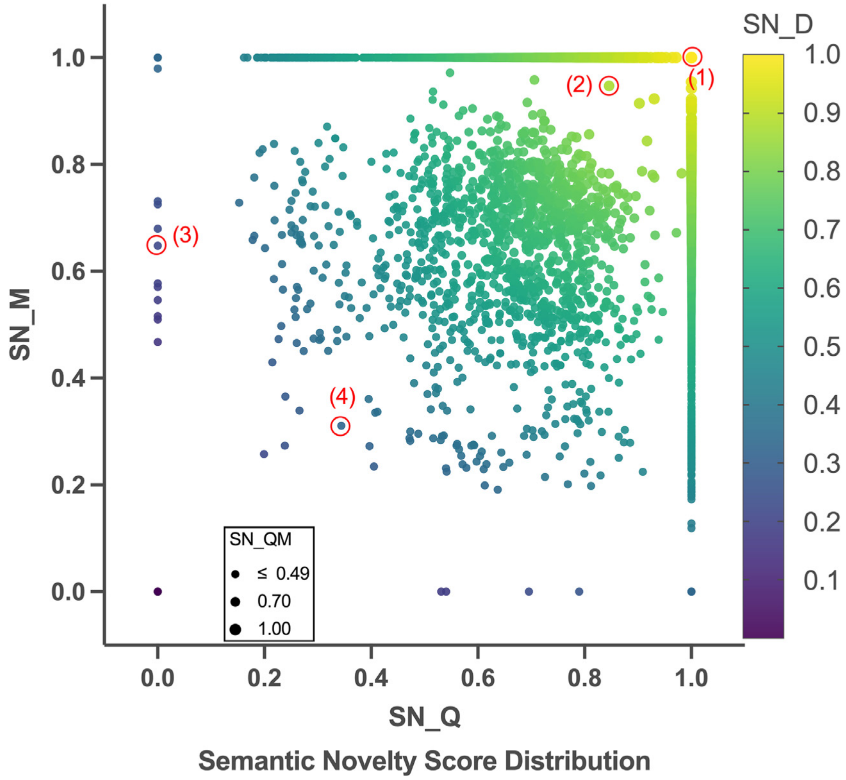

Fig. 5. Scatter plot of test papers' semantic novelty scores. The x-axis represents SN_Q, the y-axis represents SN_M, dot size (with the size of smallest symbol of 3, same as that set in Fig. 4) represents SN_QM, and dot color indicates SN_D.

is 0.491. The paper itself has the lowest novelty (0.335) of the sample papers, and the sample dot is in the lower-left corner of the graph with a darker color.

4.3. Semantic novelty measurement

After obtaining the word embedding vectors, we calculated the semantic novelty values of the test set according to Algorithm 1 and formula 6. We define four semantic novelty terms: SN_Q for semantic novelty of question term, SN_M for semantic novelty of method term, SN_QM for semantic novelty of question-method combination, and SN _D for semantic novelty of document. The descriptive statistics of the semantic index novelty scores of 12,496 articles are shown in Table 3. As shown by the statistics, the interval of the four semantic novelty indices is [0,1], and the standard deviation of SN_QM is the largest compared to LIN_QM, and its mean value is smaller, indicating that the question-method combination can capture more novelty differences. The same feature also exists between LIN_D and SN_D, indicating that the semantic novelty method can better distinguish the novelty of papers. According to the calculated scores, we produced a scatter plot of the semantic novelty score for the test data, shown in Fig. 5. Each point in the graph indicates an article.

Compared to Fig. 4, the dot distribution in Fig. 5 is more compact, and the dots with novelty value small than 1 for question or method terms appear in a clustered state. The dots in the middle are smaller compared to the dots at the upper and right edges, indicating a low score for combinational novelty. The dots with large article novelty converge toward the upper right corner, and there are a few dots near novelty of 0, which shows that there are few articles with very low novelty.

Table 4
Semantic novelty results for 4 sample articles.

id	article title	question/ first-year/ times	method/ first-year/ times	SN_Q/SN_M/ SN_QM/SN_D
1	Vibrational artificial subtle expressions: Conveying system's confidence level to users by means of smartphone vibration	artificial subtle expression/ 2018/1	smartphone vibration/ 2018/ 1	1/1/1/1
2	A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals	open data portal/ 2017/3	data govern/ 2009/ 2	0.845/0.946/ 0.833/0.875
3	Multi page search with reinforcement learning to rank	learn to rank model/ 2011/3	relevance feedback/ 1985/ 60	1.76E-16/0.679/ 1.73E-16/0.226
4	What does it mean to trust a robot?: Steps toward a multidimensional measure of trust	human-human interact/ 1998/6	trust test/ 1999/ 38	0.343/0.310/ 0.311/0.322

Table 5
Correlation test results of the two methods.

	LIN_Q vs. SN_Q	LIN_M vs. SN_M	LIN_QM vs. SN_QM	LIN_D vs. SN_D
R	0.8114	0.8157	0.1996	0.8684
R squared	0.6583	0.6654	0.03985	0.7541
P (two-tailed)	<0.0001	<0.0001	<0.0001	<0.0001
Significant? (alpha = 0.05)	Yes	Yes	Yes	Yes
Number of XY Pairs	12496			

As mentioned above, we selected four sample points for the case study, see as Table 4. Sample 1 here is the same article as the first sample in section 4.2, and the semantic novelty is 1, indicating that in this case, the two methods compute the same result. The second article investigates the question of how data governance can address the major shortcomings of government open data portals. The question term and method term appeared 3 times and 2 times, with the novelty of 0.845 and 0.946 respectively, and the article novelty was 0.875. The third article presents a feedback technique for learning to rank models, with the question term appearing 3 times, the method term 60 times, and an article novelty of 0.226. The fourth article investigates whether different aspects of trust in human-computer interaction can be considered as part of a multidimensional concept and measurement of trust. Both the question and the method arise early, and the overall novelty of the article is 0.322.

4.4. Correlation test and analysis of experimental results

To verify the consistency of the two novelty calculation methods proposed above, we conducted correlation tests on the results of the two calculations for question term, method term, question-method combination, and article novelty, respectively. We used Pearson's test for correlation analysis of the data, and the calculation results are shown in Table 5. The p-values of all four results were less than 0.0001, indicating a high significance level. In addition, in terms of R-value, the correlation coefficient for article novelty scores calculated by those methods is 0.8684, indicating a strong positive correlation. The correlations for the question term and method term were 0.8114 and 0.8157, respectively, also indicating a strong positive correlation. The correlation for the question-method combination was relatively small at 0.1996, but it is positively correlated as well.

To investigate the differences between the life-index and semantic novelty calculation methods proposed in this paper, we conduct a comparative analysis of the calculation results of the two methods. Fig. 6 shows the scores of the four metrics calculated by the two methods, with a total of eight curves. The question term novelty, method term novelty, and document novelty calculated by the two methods converge at points P2, P3 and P4 in the figure, respectively, and the trend line of semantic novelty rises more moderately in all three groups, indicating that the differences in semantic novelty calculation are more distinct and the method is better able to identify differences in term and combinational novelty. This characteristic is more evident in the question-method combination. The trend line of LIN_QM (life-index novelty of combinations) is divided into two parts, the points with novelty close to or equal to

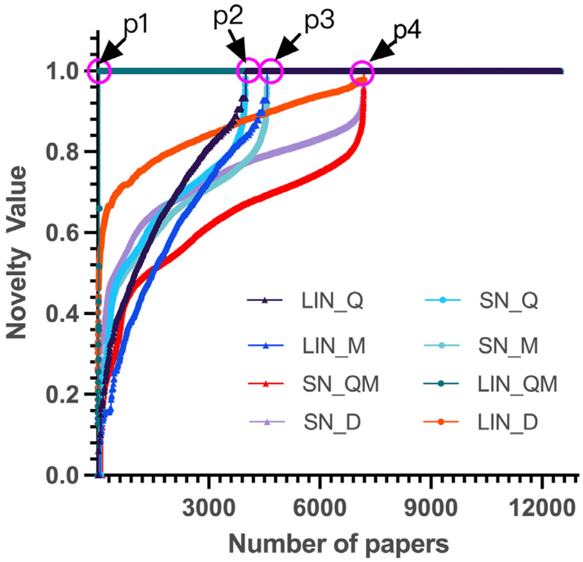

Fig. 6. Trend chart of novelty scores calculated by two methods.

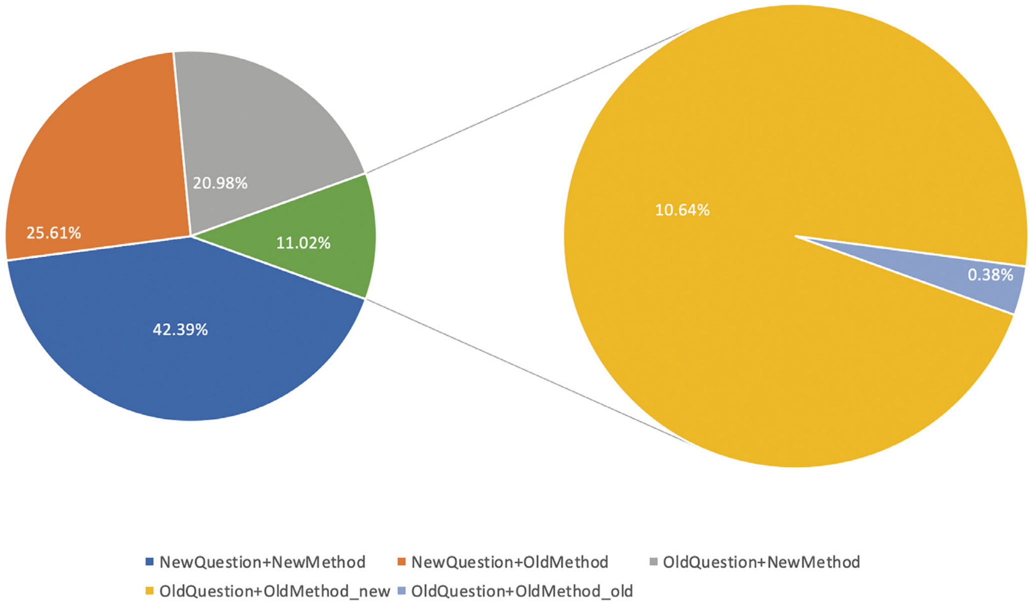

Fig. 7. Fan graph of paper novelty types.

0 are distributed near the y-axis, and the points with the remaining novelty of 1 are connected into a straight line start at point P1, indicating that the results have obvious polarization. Relatively, the trend line of SN_QM (semantic novelty of combinations) is softer and finally converges at point P4, indicating that the semantic novelty calculation approach can better express differences between combinations and thereby provide a more accurate measure of novelty.

4.5. Article novelty analysis

The novelty scores of the 12,496 articles in the test set were calculated using the two novelty methods proposed in this paper, and the statistical results showed that the median of question terms and method terms was 1, so we set the novelty threshold T_{novel} to 1. In this case, the calculation results of the two methods yielded the same proportion for the four categories of novelty classification (i.e., each proportion rounds to 25%). After further statistical analysis, we obtained the novelty type distribution of the test data as calculated by the life-index method, shown in Fig. 7. Among the four types, the largest proportion are new question + new method innovation, accounting for 42.39%, followed by new question + old method innovation (25.61%) and old question + new

Table 6
Novelty values and innovation types of randomly selected papers.

Article title	Question	Method	Citation	SN_D	LIN_D	Innovation type
MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects	deep neural network accelerators	maeri programming	156	1	1	new question + new method
Visual Domain Adaptation with Manifold Embedded Distribution Alignment	visual domain adaption	manifold embedded distribution alignment	145	1	1	new question + new method
Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec	network embedding	matrix factorization	167	0.69	0.83	new question + old method
Towards Environment Independent Device Free Human Activity Recognition	device free human activity recognition	deep learning	123	0.72	0.76	new question + old method
TEM: Tree-enhanced Embedding Model for Explainable Recommendation	person recommend	tree-enhance embed	84	0.67	0.81	old question + new method
Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate	fake news detect	structural social network	100	0.835	0.969	old question + new method
A graph-based dataset of commit history of real-world Android apps	android apps	graph-based dataset	23	0.60	0.66	old question + old method_new
Hardware Performance Counters Can Detect Malware: Myth or Fact?	malware detect	hardware performance counter	32	0.54	0.67	old question + old method_new
My Friend Leaks My Privacy: Modeling and Analyzing Privacy in Social Networks	social network	privacy protect	12	0.48	5.87E-17	old question + old method_old
Research on the Security Criteria of Hash Functions in the Blockchain	blockchain	security criteria	12	0.75	0.81	old question + old method_old

method innovation (20.98%). The combination of old question + old method is the least innovative, accounting for 11.02% of the total. Further statistics based on the life-index novelty calculation showed that 10.64% of articles used pre-existing questions and methods, but were novel in the way they were combined, being an innovative combination of existing components. Only 0.38% utilize preexisting question-method combinations, indicating that there are few such replication studies in the ACM database.

Due to the large amount of experimental data and the complexity of validation in this study, we used random sampling to randomly select 2 articles (10 articles in total) from each of the 5 innovation types mentioned above for example analysis, and the detail of the articles taken out is shown in Table 6. We obtained the citation times of articles published in 2018 in ACM database as of January 30, 2022. By analyzing the sample results, we found that combinations of new questions received more citations, and that such new questions are not completely new to the field, but are developed from existing hot topics. For example, the problem term “deep neural network accelerators” is derived from the computer science field hotspot term “deep neural network “and “visual domain adaption” is an extension of “visual domain”. In addition, new methods are emerging to study classical questions, for example, the new method "structural social network" is used to study the old question "fake news detection". The novelty of these articles has also gained high citations. For the combination of the old problem and the old method, the new combination brings more citations. As a whole, studies with novelty of question or method received more citations than articles that consisted only of old questions and old methods.

In general, the results and analyses show that among papers in computer science, few examine the same research questions using the same methods as previous studies, and the novelty of such studies is low. In addition, more than half of the problems and methods indexed in the ACM database are new, indicating that computer science is a discipline that constantly generates new research questions and methods, which also reflects its strong capacity for absorption and self-renewal.

5. Discussion

Novelty is an abstract and complex concept that is difficult to define and measure. Some studies equate the novelty of a paper with its impact and measure its novelty by the number of citations received. However, studies have shown a mixed relationship between scientific novelty and the number of citations a paper receives (Leydesdorff et al., 2019). Early studies developed the classical combinatorial innovation theory (Schumpeter & Backhaus, 2003) and applied it to novelty measurement in the research literature, which provided ideas for the novelty evaluation of scientific publications (Tahamtan & Bornmann, 2018; Uzzi et al., 2013; Wang et al., 2017). Nevertheless, we observed that 1) typical methods of novelty evaluation in the literature are based on the combination of citations, which does not reflect the essential characteristics of the innovative content; and 2) novelty measures based on keyword combinations lack semantic descriptions and may be biased against novelty.

Focusing on combinatorial novelty, we view innovative research as an innovative combination of research questions and methods. Unlike Uzzi et al. (2013), who use journal combinations to measure novelty, we build our novelty measurement from the combination of question terms and method terms in scientific publications. We extract research questions and methods from papers collected by ACM before 2018 and investigate the novelty of their question-method combinations at the time-space and semantic levels.

With continuing in-depth research on fine-grained text mining and entity extraction, word- or term-based bibliometrics is gaining more and more attention. In recent years, with the rapid development of natural language technology, semantic intelligence has once again become a hot topic, and the corresponding semantic recognition technology has also made breakthrough progress. The novelty scores derived from the life-index method are generally larger than the semantic novelty scores when the question or method words appear in the past dataset, which indicates that the life-index novelty measure is not as significant as the semantic analysis scores. The life-index only considers the frequency and time of occurrence of words or phrases, not the inner differences between individual words, and no longer works when faced with cases such as near-synonyms.

The main theoretical contribution of this paper is its introduction of a new perspective on article novelty measurement from the perspective of the question-method combination. This study distinguishes term functions and identifies the semantic roles assumed by terms in academic texts. Understanding the semantic function of author keywords in scientific and technical papers also helps to determine the accessibility and citation of papers in the scientific community (Lu et al., 2019). This combination of term function-based novelty measurement methods expands the scope of combinatorial novelty research, enables a more fine-grained characterization of paper novelty, and may further inspire cutting-edge research in innovation type identification and interdisciplinary innovation tracking.

The methodological contribution of this paper lies in its measurement of novelty from the textual content of scientific publications and its direct evaluation of innovative content (question-method terms). We not only consider the time and frequency factors of question and method terms but also include the semantic differences between terms to enrich the combined novelty, leading to two proposed novelty measurements: life-index novelty and semantic novelty. In this paper, we trained vector representations of question and method words with a powerful deep learning model to achieve a semantic-based novelty measure for question-method combinations, which provides a solution for the study of scientific publication novelty measures via deep learning techniques.

Our study has some shortcomings as well. Firstly, we identified the question words and method words of the articles based on the model proposed by Lu et al. (2020). The accuracy of the model's identification reached 83%, indicating that there are still some cases of inaccurate identification, which also affects our results to some extent. However, the data do not compromise the usability of our proposed methods, because the novelty of the articles we measure does not mean that they are innovative per se. Moreover, the novelty of in our study is not necessarily correlated with the scientific value of the article, and it is a complex issue regarding the relationship between the novelty of an article and its value or impact (Yan et al., 2020). That is to say, we focus on the "newness" of the articles, i.e., how they differ from existing studies. Secondly, we consider each paper as a combination of one question and one method; however, some papers involve more than one research question or method. For this study, the combination of core question and core method can achieve the purpose at hand, and the automatic extraction of multiple questions and methods is the next task to be solved. Thirdly, there may be disciplinary limitations. Our data come from articles in the field of computer science that have clear research questions and methods, but for humanities and social sciences disciplines such as literature and sociology, there is likely no clear question-method and our approach may be of limited applicability.

6. Conclusion

The tracking of scientific research questions and methods is important for science and technology's cutting-edge prediction and innovative research identification. Based on the combinatorial novelty theory, we focused on the term function of academic papers and proposed novelty measures for these papers based on the combination of question terms and method terms. Our study considers not only the time and frequency features of the question and method terms but also the semantic connotation of the terms themselves.

We put forward a life-index-based question-method novelty calculation method to calculate novelty scores for question terms, method terms, and question-method combinations in the test. From a semantic perspective, we proposed an algorithm for question-method combination novelty measurement based on BERT word vectors, which we used to calculate the semantic novelty of the test set articles. The results of correlation analysis and regression analysis demonstrate that the two computational methods are highly consistent with one another, indicating that the two novelty approaches proposed in this paper can be mutually supporting. Furthermore, it is found that the life-index novelty measurement method has limitations for determining differences in novelty. In comparison, the method based on semantic similarity compensates for this limitation and can measure the novelty score of words from the semantic level, which offers better differentiation and usability. Finally, based on these novelty scores, we performed a statistical analysis of the types of innovation in the test set articles. Results reflect the highly active nature of the computer science field, with the largest proportion of articles combining new problems and new methods.

This paper introduces term functions and deep learning models into the study of scientific publication novelty measurement, proposes a life-index novelty measurement and a semantic novelty calculation algorithm, which may provide new idea and method for novelty measurement study, innovation evaluation study, scientific and technological advances forecast research, etc. In a subsequent study, we intend to further explore the relationship between the novelty of scientific publications and their future impact.

CRediT authorship contribution statement

Zhuoran Luo: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Wei Lu:** Conceptualization, Data curation, Formal analysis, Writing – original draft. **Jiangen He:** Formal analysis, Writing – original draft. **Yuqi Wang:** Conceptualization, Formal analysis.

Acknowledgments

This work was supported by the Major Project of the National Social Science Foundation of China. Grant/Award Number [17ZDA292]. The authors are grateful to the anonymous referees and editors for their invaluable and insightful comments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2022.101282.

References

Amplayo, R. K., Hong, S., & Song, M. (2018). Network-based approach to detect novelty of scholarly literature. *Information Sciences, 422*, 542–557.
Arrow, K. J. (1972). Economic welfare and the allocation of resources for invention. In C. K. Rowley (Ed.), *Readings in industrial economics* (pp. 219–236). Springer.
Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8*(1), 93–102.
Chandonia, J.-M., & Brenner, S. E. (2006). The impact of structural genomics: Expectations and outcomes. *Science, 311*(5759), 347–351.
Chen, C.-H., Mayanglambam, S. D., Hsu, F.-Y., Lu, C.-Y., Lee, H.-M., & Ho, J.-M. (2011). Novelty paper recommendation using citation authority diffusion. *2011 international conference on technologies and applications of artificial intelligence* ACM Digital Library.
Chen, L., & Fang, H. (2019). An automatic method for extracting innovative ideas based on the Scopus® database. *Knowledge Organization, 46*(3), 171–186.
Cheng, Q., Li, X., & Lu, W. (2019). Construction and analysis of standard data set for domain-independent term function in academic texts. *Information Sciences, 37*(07), 41–47.
Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal, 60*(2), 433–460.
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. ***arXiv*** arXiv:1810.04805.
Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management, 37*(6), 817–842.
Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science, 47*(1), 117–132.
Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabasi, A. L. (2018). Science of science. *Science, 359*(6379), eaao0185.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.
Garfield, E. (1980). Premature discovery or delayed recognition—why? *Current Contents, 21*, 5–10.
Gupta, S., & Manning, C. D. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of 5th international joint conference on natural language processing* Asian Federation of Natural Language Processing.
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*.
Heffernan, K., & Teufel, S. (2018). Identifying problems and solutions in scientific text. *Scientometrics, 116*(2), 1367–1382.
Huang, S., & Wan, X. (2013). AKMiner: Domain-specific knowledge graph mining from academic literatures. In I*n* International Conference on Web Information Systems Engineering (pp. 241–255). Berlin, Heidelberg: Springer.
Hurd, J. M. (1992). Interdisciplinary research in the sciences: Implications for library organization. *College & Research Libraries, 53*(4), 283–297.
Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies, 76*(1), 283–317.
Kaufer, D. S., & Geisler, C. (1989). Novelty in academic writing. *Written Communication, 6*(3), 286–311.
Kondo, T., Nanba, H., Takezawa, T., & Okumura, M. (2009). Technical trend analysis by analyzing research papers' titles. In *LTC '09: Proceedings of the 4th international conference on human language technology: challenges for computer science and linguistics* (pp. 512–521). ACM Digital Library.
Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago Press 50th anniversary ed.
Lee, Y., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy, 44*(3), 684–697.
Leydesdorff, L., Bornmann, L., & Wagner, C. S. (2019). The relative influences of government funding and international collaboration on citation impact. *Journal of the Association for Information Science and Technology, 70*(2), 198–201.
Lu, W., Huang, S., Yang, J., Bu, Y., Cheng, Q., & Huang, Y. (2021). Detecting research topic trends by author-defined keyword frequency. *Information Processing & Management, 58*(4), Article 102594.
Lu, W., Li, P., Zhang, G., & Cheng, Q. (2020). Recognition of lexical functions in academic texts: Automatic classification of keywords based on BERT vectorization. *Journal of the China Society for Scientific and Technical Information, 39*(12), 1320–1329.
Lu, W., Li, X., Liu, Z., & Cheng, Q. (2019). How do author-selected keywords function semantically in scientific manuscripts? *Knowledge Organization, 46*(6), 403–418.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. ***arXiv preprint*** arXiv:1301.3781.
Min, C., Bu, Y., & Sun, J. (2021). Predicting scientific breakthroughs based on knowledge structure variations. *Technological Forecasting and Social Change, 164*, Article 120502.
Min, C., Bu, Y., Sun, J., & Ding, Y. (2018). Is scientific novelty reflected in citation patterns? *Proceedings of the Association for Information Science and Technology, 55*(1), 875–876.
Mukherjee, S., Uzzi, B., Jones, B., & Stringer, M. (2016). A new method for identifying recombinations of existing knowledge associated with high-impact innovation. *Journal of Product Innovation Management, 33*(2).
Nanba, H., Kondo, T., & Takezawa, T. (2010). Automatic creation of a technical trend map from research papers and patents. In *Proceedings of the 3rd international workshop on patent information retrieval* ACM Digital Library.
Nelson, R., & Winter, S. (1982). *An evolutionary theory of economic change*. Harvard University Press.
Packalen, M., & Bhattacharya, J. (2015). Age and the trying out of new ideas (No. w20920). *National Bureau of Economic Research*.
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)*. Association for Computational Linguistics.
Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, 1*.
Pimentel, M. A. F., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing, 99*, 215–249.
Ponomarev, I. V., Williams, D. E., Hackett, C. J., Schnell, J. D., & Haak, L. L. (2014). Predicting highly cited papers: A method for early detection of candidate breakthroughs. *Technological Forecasting and Social Change, 81*, 49–55.
Ruibin, W. (2018). Empirical research on the topic innovation of the paper based on self-citation network and main path analysis. *Library and Information Service, 62*(03), 64–70.
Schumpeter, J., & Backhaus, U. (2003). The theory of economic development. In *Joseph Alois Schumpeter* (pp. 61–116). Springer.
Schumpeter, J. A. (1939). Business cycles. New York: Mcgraw-hill.
Shibata, N., Kajikawa, Y., & Matsushima, K. (2007). Topological analysis of citation networks to discover the future core articles. *Journal of the Association for Information Science and Technology, 58*(6), 872–882.

Shibata, N., Kajikawa, Y., Takeda, Y., Sakata, I., & Matsushima, K. (2010). Early detection of innovations from citation networks. *IEEE international conference on industrial engineering & engineering management*.

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics, 68*(3), 595–610.

Soler, J. M. (2007). A Rational indicator of scientific creativity. *Journal of Informetrics, 1*(2), 123–130.

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy, 44*(8), 1445–1461.

Tahamtan, I., & Bornmann, L. (2018). Creativity in science and the link to cited references: Is the creative potential of papers reflected in their cited references? *Journal of Informetrics, 12*(3), 906–930.

Tang, R. (2004). Evolution of the interdisciplinary characteristics of information and library science. *Proceedings of the American Society for information Science and Technology, 41*(1), 54–63.

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. J., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science, 342*(6157), 468–472.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American journal of sociology, 111*(2), 447–504.

Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics, 59*(3), 467–472.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA.

Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy, 45*(3), 707–723.

Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C.-C. J. (2019). Evaluating word embedding models: Methods and experimental results. *APSIPA Transactions on Signal and Information Processing, 8*.

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.

Weis, J. W., & Jacobson, J. M. (2021). Learning on knowledge graph dynamics provides an early warning of impactful research. *Nature Biotechnology, 39*(10), 1300–1307.

Yan, Y., Tian, S., & Zhang, J. (2020). The impact of a paper's new combinations and new components on its citation. *Scientometrics, 122*(2), 895–913.

Zhang, J., Yan, Y., & Guan, J. (2019). Recombinant distance, network governance and recombinant innovation. *Technological Forecasting and Social Change, 143*, 260–272.