Journal of King Saud University – Computer and Information Sciences 35 (2023) 101569

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Detecting interdisciplinary semantic drift for knowledge organization based on normal cloud model

Zhongyi Wang [a], Siyuan Peng [a], Jiangping Chen [b,c], Amoni G. Kapasule [d], Haihua Chen [b,c,*]

[a] School of Information Management, Central China Normal University, Wuhan 430079, China
[b] Department of Information Science, University of North Texas, Denton 76203, TX, USA
[c] Intelligent Information Access Lab, University of North Texas, Denton 76203, TX, USA
[d] Kamuzu University of Health Sciences, Lilongwe, Malawi

A R T I C L E I N F O

Article history:
Received 7 February 2023
Revised 19 April 2023
Accepted 21 April 2023
Available online 27 April 2023

Keywords:
Interdisciplinary semantic drift
Knowledge representation
Normal cloud model
Knowledge potential energy
Knowledge potential difference

A B S T R A C T

To reduce the conceptual ambiguity in interdisciplinary knowledge organization systems (KOSs) and enhance interdisciplinary KOS management, this paper proposes a framework for interdisciplinary semantic drift (ISD) detection based on the normal cloud model (NCM). In this framework, we first analyze the features of interdisciplinary concepts and propose a novel interdisciplinary concept extraction method based on cross-discipline statistical information. Secondly, the high-performance knowledge representation model NCM is adopted to represent each interdisciplinary concept with uncertainty, and then a new ISD degree calculation method is proposed based on the similarity cloud algorithm. Thirdly, to identify the direction of ISD after the degree calculation, we propose an ISD direction identification method according to the theory of knowledge potential energy (KPE). Fourthly, based on the above procedure, we propose an ISD detection algorithm to identify and visualize the ISD process. Finally, we evaluate the proposed framework on the concept of "information entropy" and compare the performance with three baselines. Experimental results demonstrate that our framework outperforms[all the baselines, and the result is comparable to experts' judgments (0.808 on Spearman correlation, $p<0.001$). The research indicates the meaning of an interdisciplinary concept will drift from the high KPE discipline to the low KPE discipline as long as interdisciplinary knowledge potential differences (KPD) exist between these two related disciplines. We further identify three key factors that affect the degree of ISD: the length of the discipline chain of an interdisciplinary concept transfer, the number of source disciplines that an interdisciplinary concept comes from, and the knowledge distance between the source discipline and the target discipline.

1. Introduction

Contending with increasingly complex challenges, such as climate change, food and water crises, and public health emergencies that our societies face, interdisciplinary research is attracting more and more attention because of its capacity to provide knowledge for handling these complicated dilemmas (Feldhoff et al., 2019). In this situation, researchers often seek innovative solutions to these hardships in related studies in other disciplines. Nevertheless, they usually must familiarize themselves with other research fields, otherwise, they will face significant hurdles in interdisciplinary knowledge access and dissemination. Therefore, knowledge organization for helping researchers to discover related literature from other disciplines is an absolute necessity (He et al., 2022). Although there are various knowledge organization systems (KOSs), including classification schemes, subject heading lists, thesauri, and ontologies (Browne et al., 2019; de Almeida and Farinelli, 2017), they are generally only constructed for particular disciplines of study. They poorly serve interdisciplinarity. In other words, these KOSs are helpful for people to access scientific literature within a discipline, but they need to be modified to make them easier for researchers to identify related scientific literature

* Corresponding author at: Department of Information Science, University of North Texas, Denton 76203, TX, USA.

E-mail addresses: wzywzy13579@163.com (Z. Wang), siyuan.peng@mails.ccnu.edu.cn (S. Peng), Jiangping.chen@unt.edu (J. Chen), akapasule@kuhes.ac.mw (A.G. Kapasule), haihua.chen@unt.edu (H. Chen).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2023.101569
1319-1578/

in other disciplines. While more and more information scientists recognize the increased importance of interdisciplinarity, they also recognize the obstacles it poses to interdisciplinary knowledge organization (Greenberg et al., 2021). As research and knowledge become more interdisciplinary, an interdisciplinary KOS that is easy to understand and does not place arbitrary barriers between disciplines or between the academy and societal actors is needed. With the number of scientific literature published today vastly increasing, researchers are more reliant on interdisciplinary KOSs because they are in the nature of interdisciplinarity to seek relevant insights from multiple bodies of knowledge.

Specifically, interdisciplinary KOSs should allow users to search for particular theories, methods, and other related knowledge on a phenomenon from different disciplines (Szostak et al., 2016). Therefore, the goal of interdisciplinary KOSs would be a general classification that respects the concepts employed in each discipline. In this situation, concepts become one of the main problems when building interdisciplinary KOSs. Most importantly, disciplines often borrow concepts from other disciplines, which are called interdisciplinary concepts in this paper. In order to match different disciplinary contexts, interdisciplinary concepts often acquire new meanings when new disciplines adopt them. This is especially true in computer science and engineering, as it is common for new concepts to be derived from other domains so that they can adequately illustrate the novelty of the published research works in a concise yet accurate manner (Paharia et al., 2021). Therefore, before constructing an interdisciplinary KOS, the first question that must be asked is whether each interdisciplinary concept carries the same meaning in different disciplines. For example, a term such as 'abortion' carries a pretty different connotation in gender studies than in medicine. In such cases, we should first identify the degree of differences in definition across disciplines, called interdisciplinary semantic drift (ISD) (Stavropoulos et al., 2019; Kutuzov et al., 2018). For example, a concept refers to different objects in different disciplines, or a concept has very different properties in various domains. This articulates the need to detect, measure, and interpret ISDs of concepts among different disciplines. In other words, the semantic drifts of concepts across disciplines should be investigated. If an interdisciplinary KOS fails to detect ISDs, it will be characterized by concept ambiguity. Special care must be taken to reduce the ambiguity that exists for a large number of concepts.

Different fields (e.g., ontology management and versioning) are studying semantic drift in KOSs, and various methods have been applied. Semantic drift research is beneficial for the multitude of domains with KOS-based applications. Even though ontology management and versioning tools are now available, they are of limited use for ontology evolution unless the desired changes are known beforehand. The reason is that the process of ontology evolution consists of change capturing, change representation, the semantics of change, change implementation, change propagation, and change validation (Wardhana et al., 2018). While ontology management and versioning mainly deal with the representation, implementation, and propagation of changes, the more difficult part of semantic drift capturing has been left to ontology learning approaches. These methods explore machine learning techniques to capture semantic drifts based on similarity measures. According to the aspects considered in semantic drift assessment, these methods can be further divided into element-based, structure-based, and background-based methods. Element-based methods detect semantic drift by computing the similarity of concepts' elements (such as instances, labels, and others). For example, Stavropoulos et al. (2019) combined label-based and morphing-based approaches to detect semantic drifts of concepts. Structure-based methods exploit the structure of the ontologies to determine the semantic drift of concepts in two ontologies. For example, Capobianco et al. (2020) proposed an approach to detect and assess semantic drift by inspecting the taxonomic structure among concepts. However, both element- and structure-based methods capture the semantic drift using only information gathered from the input ontologies. They cannot capture semantic drifts of concepts without any lexical or structural similarity. In this situation, concept matching with background knowledge is more suitable for semantic drift capturing (Husein et al., 2016). Unlike the above methods that limit the use of information only provided by the input ontologies, background-based methods utilize external resources to help to detect semantic drifts of concepts (Annane et al., 2018). This method has become prevalent as it captures semantic drifts for concepts without lexical or structural similarity.

Concerning previous research in measuring diachronic semantic drift, this paper extends background-based methods to measure ISD. Specifically, this paper aims to detect semantic drifts in interdisciplinary concepts. To achieve the research goal, we develop the following research questions:

- **What data should be selected as the background knowledge?** The first challenge for using the background-based method is the appropriate use of background knowledge (Shvaiko and Euzenat, 2013). What data is used as background knowledge affects not only the interdisciplinary concept representation but also the ISD detection (Tahmasebi et al., 2018). Therefore, background-based interdisciplinary semantic drift capturing must be done on high-quality datasets representative of the concepts used in different disciplines. Specifically, the background-based method for ISD capturing is more applicable if there are rich descriptions of concepts and the descriptions are constantly modified over disciplines. Compared to newspapers (Yao et al., 2018), Wikipedia (Takamura et al., 2017), etc., which are commonly used for semantic drift detection, the scientific literature, with sufficient volume and disciplinary span, is more capable of meeting the above two conditions; therefore, we take it as the background knowledge of concepts to identify their ISDs. Analyzing textual descriptions of scientific literature can tell us how researchers deal with an interdisciplinary concept in different disciplines.
- **How to extract interdisciplinary concepts from scientific literature?** In the interdisciplinary KOSs, interdisciplinary concepts are the central constructs used to describe sets of objects with shared characteristics. Automatic term extraction, an automated process of identifying terms (linguistic representations of concepts) in texts, is beneficial to interdisciplinary concepts extraction (Hazem et al., 2020). However, despite being a well-established research domain for decades, automatic term extraction still fails to meet human standards (Lang et al., 2021), especially for interdisciplinary concepts extraction. Therefore, identifying interdisciplinary concepts from text is still a challenge. To deal with this problem, this paper first analyzes the features of interdisciplinary concepts. We find that different from the concepts in a thematically restricted discipline, interdisciplinary concepts are domain-specific terms that are also used by other disciplines. Based on the above analysis, this paper proposes an effective interdisciplinary concepts extraction method based on cross-discipline statistical information, which can automatically extract interdisciplinary concepts from the scientific literature. Specifically, this method includes three steps: candidate term extraction, domain-specific term selection, and interdisciplinary term selection.
- **How to represent interdisciplinary concepts effectively?** The key challenge to background-based methods is selecting a proper text representation for background knowledge sources

when implementing background-based semantic drift capturing. To solve this problem, a variety of techniques, ranging from counting approaches over generative models to neural network-based word embedding, such as a random indexing algorithm (Basile et al., 2014), SVD, SGNS (Hamilton et al., 2016), Word2Vec (Yao et al., 2018), and BERT (Laicher et al., 2021), are used to model word meaning directly. However, concepts are usually designated by terms. One of the characteristics of conceptual terms is uncertainty, such as fuzziness and randomness. For example, it is difficult to obtain exact intentions or establish precise boundaries of the extensions of concepts such as "cold", "beautiful", and other related concepts. However, the above knowledge representation models cannot formally describe the uncertainties of interdisciplinary concepts effectively. However, the normal cloud model (NCM) can synthetically describe the randomness and fuzziness of concepts (Wang et al., 2021), so we adopt the high-performance knowledge representation model NCM for ISD detection.

- **How to calculate the semantic drift degree of interdisciplinary concepts?** The degree of ISDs can be calculated by comparing concept representations between two or more disciplines. Different concept representation models entail different mathematical functions to quantify the change. For example, cosine or Euclidean distances are often used for VSM, Word2Vec, and BERT (Hengchen et al., 2021). However, different from these concept representation methods, this paper uses the cloud model to represent interdisciplinary concepts; therefore, similarity cloud algorithm (Yao et al., 2021) is applied to calculate the semantic drift degree of interdisciplinary concepts accordingly. To evaluate the method, we conduct an empirical study on the interdisciplinary concept "information entropy" using the China National Knowledge Infrastructure (CNKI) academic literature dataset and then compare our proposed approach with VSM, Word2Vec, and BERT. The experimental results show that our proposed NCM-based approach achieves the best performance, is highly coincident with the experts' judgment results, and that the correlation is significant.
- **How to identify the semantic drift direction of interdisciplinary concepts?** To help people understand the ISDs, we should also portray the process of ISDs of interdisciplinary concepts among different disciplines. Related to this is the need for more extensive visualization approaches for ISDs. Therefore, the direction of ISD should also be identified after calculating the degree of ISD. To solve this problem, based on the theory of knowledge potential energy (KPE) (Hai et al., 2007), this paper proposes an approach to identify the direction of ISD by citation analysis (Zeng et al., 2021). We believe that as long as there is interdisciplinary knowledge potential difference (KPD) between related disciplines, the concept will drift from the high-KPE discipline to the low-KPE discipline. Specifically, we detect the direction of ISD between disciplines by monitoring cross-disciplinary citations. Finally, we visualize the semantic drift degree and direction of each interdisciplinary concept to support interdisciplinary KOSs' construction and application.

The rest of this article is organized as follows: Section 2 reviews the related work regarding existing approaches to ISD measurement. Section 3 describes the methodology of ISD detection and visualization. Section 4 presents the experiments and analyzes the results. Further analysis and implications are discussed in Section 5. Finally, Section 6 concludes the article and discusses future work.

2. Related work

With the development of Semantic Web technologies, semantic drift is becoming an active area of research. Considering the potential to significantly impact and benefit the multitude of KOSs-based applications, more and more attention is being paid to semantic drift. However, capturing semantic drift is a complex task. In addition, semantic drift relates to various lines of research, such as KOS evolution, change, management, and versioning (Meroño-Peñuela et al., 2013). Methods of semantic drift capturing are mainly performed based on similarity measures. These kinds of methods explore machine learning techniques to capture semantic drift. According to the aspects considered in semantic drift assessment, these methods can be further divided into element-based, structure-based, and background-based methods.

2.1. Element-based methods

Element-based methods detect semantic drift by computing the similarity of concepts' elements (such as labels, instances, etc.) in the two input KOSs. Label-based methods calculate the similarity score between two strings. They are widely used techniques in semantic drift detecting and can be categorized into the following families: edit-based similarity and token-based similarity (Cohen et al., 2003). Instance-based methods exploit the extension of the concepts in the KOSs. With the intuition that if the instances are alike, the concepts they belong to should be similar. These methods capture the semantic drift of concepts by analyzing their instances. The most frequently used metrics are Jaccard similarity and k-Statistic. Recently, Stavropoulos et al. (2019) proposed a hybrid method for semantic drift detection, which combines label-based and morphing-based approaches.

2.2. Structure-based methods

Structure-based methods exploit the structure of the KOSs to determine the semantic drift of concepts in two KOSs. These methods are based on the structure of the concepts found in the KOS. The internal structure consists of the entity name, annotations, properties, and relations to other entities. A KOS is considered a graph or tree structure in structure-based techniques. The similarity between the two concepts is based on their position in the graph. If two concepts are similar, their neighbor concepts must have some similarity (Khan and Safyan, 2014). Thus, finding the semantic drift of concepts between KOSs is related to finding a maximum common directed subgraph. For example, Eder and Wiggisser (2007) proposed semantic drift detection of an ontology based on the directed acyclic graph. The approach compares versions based on the graph structure representing an ontological structure.

2.3. Background-based methods

Unlike the above methods that limit the use of information only provided by the input KOSs, background-based methods utilize external resources to help to detect semantic drifts of concepts (Chen, 2014). A background could provide descriptions and definitions that can be used in detecting semantic drifts of concepts. Many linguistic resources, such as lexicons, catalog metadata, and web pages, can be exploited and used as references for the alignment process. For example, WordNet, an English lexical database, has often been used for measuring the similarity of concepts (Cross et al., 2013). Catalog metadata was employed to detect semantic drift based on a machine learning clustering algorithm ESOM to express the evolving semantics of concepts (Darányi et al., 2016). WikiMatch (Hertling and Paulheim, 2012) uses Wikipedia as its background knowledge. Their experiments showed that Wikipedia could help improve the results in terms of recall and precision. Web pages, which include the text describing concepts, were used by Gulla et al. (2010) to detect semantic drift in KOSs.

The availability of large background knowledge sets has enabled the development of new methodologies for studying semantic drift. However, one challenge to background-based methods is the appropriate use of background knowledge (Shvaiko and Euzenat, 2013). When implementing background-based semantic drift capturing, selecting a proper text representation for background knowledge sources may be crucial. To model word meaning more directly, the vector space model (VSM) is used to solve this task. The assumption of VSM is that words occurring in similar contexts tend to have similar meanings. Although VSM can imitate the field-like continuity of conceptual content, vector space in its most basic form is not semantic (Darányi et al., 2016). For example, Gulla et al. (2010) proposed a vector-based approach to detect semantic drift in ontologies. The method uses concept signatures to represent a concept construct based on how concepts are described. However, the signature is not a semantic representation of a concept. It merely shows how words and linguistic expressions refer to and discuss the concept. Therefore, distributional word representations were provided to solve this task. They represent meaning with sparse or dense (embedding) vectors produced from word co-occurrence counts. A fundamental assumption of these methods is that changes in a word's collocational patterns reflect changes in word meaning. For example, Wittek et al. (2015) employed the random indexing algorithm to create word vectors. Sagi et al. (2011) turned to latent semantic analysis. Kutuzov et al. (2018) found that these distributional methods seemed well suited for monitoring the gradual process of meaning change. Recently, word embeddings, like GloVe (Pennington et al., 2014) or Word2Vec (Yao 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. For example, Kim et al. (2014) provided a method for detecting semantic drift through a chronologically trained neural language model, which was trained on the Google Books Ngram corpus. However, the above representation models belong to static word embedding since they only produce a fixed context-independent representation for each word and do not address polysemy during concept representation. To solve this problem, BERT, as contextual word embedding, produces the word-level representation based on the information of the entire sentence and thus can obtain much better results in the concept representation task. Following its widespread adoption in NLP in general (Li et al., 2020), it has become the dominant representation for the analysis of diachronic semantic drifts as well (Giulianelli et al., 2020). However, language is nebulous, which leads to the uncertainty in the expression of concepts. BERT is often influenced by uncertainty, especially the variety of the target concept forms, which makes BERT unable to generate high-quality vector representations for concepts (Gupta et al., 2019), causing its low performance on semantic drift detection (Hengchen et al., 2021).

However, linguists have long recognized that language is nebulous (Tahmasebi et al., 2018) because concepts are continuous while their mapping to language is discrete, which leads to uncertainty in the expression of concepts. The uncertainties of concepts present a series of challenges for identifying their ISDs. Among the uncertainties involved in concepts, randomness and fuzziness are the two most essential uncertainties which have attracted much attention. The fuzziness of an interdisciplinary concept mainly refers to the uncertainty about the range of extension of the concept. The randomness of concepts means that any concept is not an isolated fact but is related to the external world in various ways. Moreover, the fuzziness and randomness of a concept are usually tightly related and inseparable. However, the existing uncertain knowledge representation models based on probability theory and fuzzy sets theory cannot formally describe the inherent relation between randomness and fuzziness and integrate both uncertainties effectively. On the other hand, NCM can synthetically describe the randomness and fuzziness of concepts (Wang et al., 2021). To this end, we employ the high-performance knowledge representation model NCM for ISD identification.

In conclusion, initial methods, such as element- and structure-based methods, capture the semantic drift using only information gathered from the input KOSs. One of the sources of difficulty for matching tasks is that KOSs are designed with certain background knowledge and in a specific context, which unfortunately does not become part of the KOS specification, and, thus, are unavailable to match terms. The lack of background knowledge increases the difficulty of the matching task by generating too many ambiguities. Additionally, they cannot capture semantic drifts of concepts without lexical or structural similarity. Concept matching with background knowledge is more suitable for semantic drift capturing in this situation. Therefore, when there is not any lexical or structural similarity between concepts, using background knowledge in the similarity measure is reasonable as it can capture concepts' semantic drifts effectively in these cases. To deal with the uncertainties involved in concepts, we employ NCM to represent each interdisciplinary concept during the ISD detection.

3. Methodology

In this section, we propose a novel framework for ISD detection based on the NCM (see Fig. 1). The framework includes four steps: (1) Interdisciplinary concept extraction, (2) Interdisciplinary semantic drift degree calculation, (3) Interdisciplinary semantic drift direction identification, and (4) Interdisciplinary semantic drift detection algorithm. In the first step, we provide a new method to extract interdisciplinary concepts from scientific

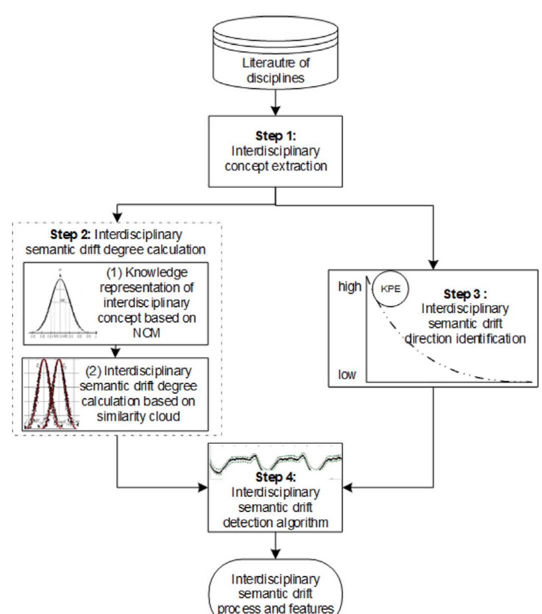

Fig. 1. The framework for interdisciplinary semantic drift detection based on the normal cloud model.

literature based on cross-discipline statistical information. Then in the second step, we calculate the ISD degree of the extracted interdisciplinary concepts based on the NCM. In the third step, we detect the ISD directions of each interdisciplinary concept between disciplines according to the theory of KPE. In the last step, we propose an algorithm of ISD detection based on the above procedure. We will describe the detailed implementation of these four steps in the following subsections.

3.1. Interdisciplinary concept extraction

A concept can be a single term or a combination of multiple terms defined as the lexical units used to represent the meaning of the concept in a thematically restricted discipline. Unlike the concepts in a thematically restricted discipline, interdisciplinary terms are domain-specific terms also used by other disciplines. However, identifying these interdisciplinary terms is challenging. Interdisciplinary terms usually have higher domain relevance and are distributed unevenly among disciplines compared to common words. In other words, interdisciplinary terms usually appear with a higher frequency in the initial domain while having a lower frequency in other domains or may not appear at all in other domains. Inspired by this finding, this paper proposes an interdisciplinary term extraction method by eliminating common words from a group of documents. The workflow of the proposed interdisciplinary concept extraction approach is presented in Fig. 2. The process includes three steps: candidate term extraction, domain-specific term selection, and interdisciplinary term selection. The core idea is to measure the domain relevance of keywords.

3.1.1. Candidate term extraction

To extract high-quality term candidates, existing studies experimented with various degrees of linguistic filtering (Wong et al., 2007). However, none of the existing approaches is satisfactory for practical applications, so an alternative strategy is necessary and urgent. Instead of extracting the terms from an article directly, in this research, we use the keywords assigned by authors as candidate terms since these keywords represent the main topics of the article.

3.1.2. Domain-specific term selection

Existing approaches of extracting domain-specific terms from the candidate terms mainly focus on computing their term hood, meaning the degree a candidate qualifies as a terminological unit (Basili et al., 2001; Hätty et al., 2017; Wong et al., 2007). However, the drawback to these approaches is that they only rely on the distributional information of a single target discipline domain. We propose a domain-specific term selection strategy based on cross-discipline statistical information to overcome this drawback. Specifically, we leverage the idea of the term frequency-inverse document frequency (tf-idf) method to calculate the term hood $Th(k_i)$ of a candidate term k_i, which is presented in Eq. (1).

$$Th(k_i) = \frac{n_{i,j}}{\sum_{j}^{m} n_{i,j}} \times \log\frac{m}{Id_i} \tag{1}$$

where m is the total number of disciplines, $n_{i,j}$ is the frequency of the term i in the discipline j, and Id_i is the number of disciplines in which that term i occurs. Candidate terms whose $Th(k_i)$ scores are higher than the threshold γ are selected as terms while others are removed.

3.1.3. Interdisciplinary term selection

Intuitively, terms with a high level of interdisciplinarity should appear in many disciplines. Based on this idea, interdisciplinary terms should also have a high degree of discipline distribution (Thorleuchter and Van den Poel, 2016). In other words, the more diverse the disciplines in which a term occurs, the more likely this term becomes an interdisciplinary term. We measure the disciplinary diversity of a term based on Eqs. (2) and (3).

$$E(k_i) = \frac{H(k_i)}{logm} \tag{2}$$

$$H(k_i) = -\sum_{j=1}^{m} (p(k_{ij}) \times \log p(k_{ij})) \tag{3}$$

where $p(k_{ij})$ is the occurrence probability of term i in discipline j, m is the total number of all disciplines.In this study, terms whose $E(k_i)$ score is higher than the threshold λ are selected as the interdisciplinary terms while others are removed.

3.2. Interdisciplinary semantic drift degree calculation

Before calculating the degree of ISD, we need to clarify the connotation of ISD first. In this research, we define "ISD" as follows:

Definition 1: Interdisciplinary semantic drift (Stavropoulos et al., 2016). The phenomenon of change in concept meaning over different disciplines. The semantic of an interdisciplinary concept C has drifted between discipline i and j if and only if the similarity degree $sim(C_i, C_j)$ is less than a threshold. Otherwise, if an interdisciplinary concept in different disciplines has the same meaning, there is no ISD.

From the definition of ISD, we can see that the key to obtaining the degree of ISD is calculating the semantic distance of an interdisciplinary concept between different disciplines. However, the uncertainties of concepts present a series of challenges for the detection of ISD. For example, how to describe qualitative interdisciplinary concepts formally in natural language? To solve this problem, we propose using the NCM to represent a qualitative concept quantitatively, then calculate the ISD degree of the interdisciplinary concept based on its knowledge representation model in different disciplines.

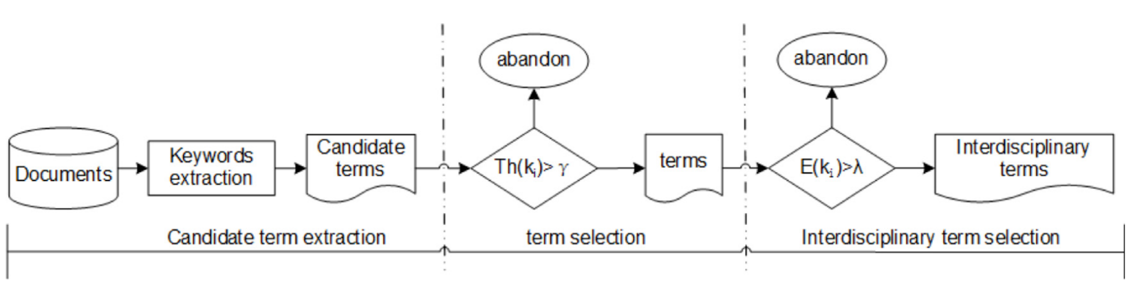

Fig. 2. Interdisciplinary concept extraction from the scientific literature.

3.2.1. Knowledge representation of an interdisciplinary concept based on NCM

An interdisciplinary concept, as a unique concept, usually has a higher uncertainty than a general one since it might have different meanings under different scenarios. Therefore, accurately modeling the uncertainties of interdisciplinary concepts becomes a fundamental task of this research. Unlike the existing studies based on probability or fuzzy set theory, we propose a quantifiable interdisciplinary concept knowledge representation approach based on the NCM.

(1) Cloud model.

There are different cloud models, such as normal cloud and γ cloud. Since most uncertainty concepts behave as normal clouds in nature, we select the NCM instead of the γ cloud to represent the uncertainties of interdisciplinary concepts. The cloud model is defined as the following:

Definition 2: Cloud and cloud drops (Qin et al., 2011**).** 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, and $\mu(x) \in [0, 1]$, standing for the certainty degree for which x belongs to c, is a random variable with stable tendency, then:

$$\mu : U \to [0, 1]\ \forall x \in U\ x \to \mu(x) \tag{4}$$

The distribution of x in the universe of discourse U is called the cloud, and each x is called a cloud drop. The character of the concept is expressed through all the cloud drops. The certainty degree of a cloud drop can be understood as the extent to which the drop can represent the concept accurately.

Definition 3: Normal cloud and cloud drops (Ma and Zhang, 2020**).** Let U be the universe of discourse and c be a qualitative concept in U. If $x \in U$ is a random instantiation of concept c which satisfies $K_j \sim N(Ex,\ En'^2), En' \sim N(En,\ He^2)$, and the certainty degree of x belonging to concept c satisfies Eq. (5):

$$\mu(x) = e^{-\frac{(x-Ex)^2}{2(En')^2}} \tag{5}$$

Then the distribution of x in the universe U is called a normal cloud, and every $(x, \mu(x))$ is defined as a cloud drop.

We use numerical characteristics, including expected value (Ex), entropy (En), and hyper entropy (He) to represent the property of a concept in the NCM. Ex is the mathematical expectation of the cloud drops distributed in the universal set. In other words, the element Ex is the position in the universe of discourse, corresponding to the center of gravity of the cloud. En is the fuzziness measurement of the qualitative concept showing how many elements in the universe of discourse could be accepted as the linguistic atom. He is the randomness measurement of the entropy En. In the NCM, drops that have significant contributions to the qualitative concept are mostly in the range: $[Ex - 3En, Ex + 3En]$. The other cloud drops located outside the range are called the small probability event. The overall characteristics of the NCM are not affected if we do not consider them. This is the "3En" rule of a normal cloud. Fig. 3 shows an example of the interdisciplinary concept of a normal form cloud whose numerical characteristics are $Ex = 0, En = 3$, and He is 0.3.

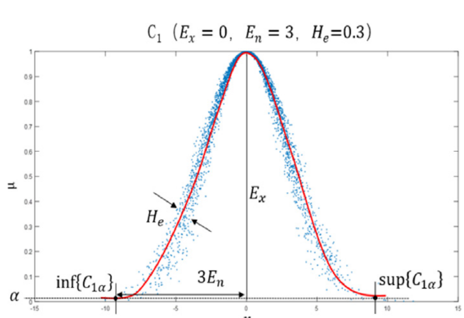

Fig. 3. The normal form cloud of an interdisciplinary concept.

(2) NCM for an interdisciplinary concept.

In this paper, to represent interdisciplinary concepts based on the NCM, we first give definitions correlated with a normal cloud of an interdisciplinary concept as follows.

Through deep analysis, we find that if the meaning of an interdisciplinary concept is stable in different disciplines (the connotation has not changed), then the concepts related to it (that is, the concepts that co-occur with it) in different disciplines are also relatively unchanged or stable. If an interdisciplinary concept is unstable, the related concepts will also be unstable. In other words, semantic drifts are often reflected in large corpora through changes in the context of the word, which is undergoing a shift, as measured by co-occurring words. This phenomenon aligns well with the assumptions that changes in a word's collocational patterns reflect changes in word meaning (Hilpert, 2008). Kutuzov et al. (2018) found that these distributional methods seem well suited for monitoring the gradual process of semantic drift. Therefore, we use co-occurred keywords to detect the ISD as the cloud drops of an interdisciplinary concept. The cloud drops for the interdisciplinary concept are described in Definition 4.

Definition 4: Cloud drops for an interdisciplinary concept (Li et al., 2009**).** Each cloud drop is a tuple $(k_j, \mu_i(k_j))$. k_j is the j^{th} keyword which co-occurs with interdisciplinary concept c_i. $\mu_i(k_j)$ is the degree of similarity (DoS) between the interdisciplinary concept c_i and its related keywords k_j. Therefore, the cloud drops of the interdisciplinary concept c_i can be represented as drops $(k_j, \mu_i(k_j)) = ((k_1, DoS_{i1}), (k_2, DoS_{i2}), \ldots (k_j, DoS_{ij}), and \ldots (k_n, DoS_{in}))$. The function of DoS is described in Eq. (6). $DoS \in [0, 1]$. 0 and 1 represent the lower and upper limits of the similarity.

$$DoS_{ij} = \frac{f_{i,j}}{\sqrt{f_i \times f_j}} \tag{6}$$

where f_i is the occurrence frequency of interdisciplinary concept c_i in the set of documents, f_j is the occurrence frequency of related keywords k_i in the set of documents, and f_{ij} is the co-occurrence frequency of c_i and k_i.

Definition 5: Cloud generator (CG) (Yan et al., 2019**).** CG is the intermediate converter between qualitative descriptions and quantitative calculations and includes two cloud generators: forward cloud generator and backward cloud generator.

To achieve the knowledge representation of an interdisciplinary concept based on an NCM, the reverse normal cloud generator is used to transform the cloud drops describing quantitative values into the qualitative vector features Ex, En, and He. The vector $C(Ex, En, He)$ is called a cloud vector representing a qualitative concept.

Definition 6: Backward normal cloud generator (BNCG) (yu Zhang et al., 2016**).** The BNCG can be used to calculate the numerical characteristics of the sample cloud on the base of the cloud drops formed by the sample, as shown in Fig. 4.

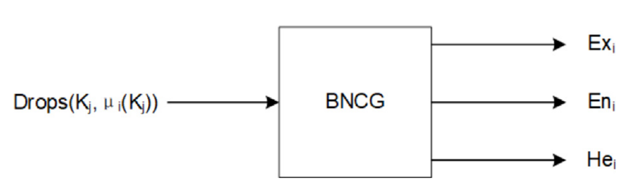

Fig. 4. Backward normal cloud generator.

Through BNCG, the digital characteristics (Ex, En, He) of an interdisciplinary concept normal cloud can be obtained based on the cloud drops of the interdisciplinary concept. The procedures for calculating digital characteristics based on BNCG are as follows (see Algorithm 1).

Algorithm 1 Generation process of a cloud vector based on the BNCG.

Input: $Drops(k_j, DoS_{ij}), j = 1, 2, \ldots n.$
Calculate the mean of DoS_{ij} : $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} DoS_{ij}$
Calculate the variance of DoS_{ij} : $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(DoS_{ij} - \bar{X}_i)^2$
Obtain the expectation: $Ex_i = \bar{X}_i$
Calculate entropy: $En_i = \sqrt{\frac{\pi}{2}} \times \frac{1}{n}\sum_{j=1}^{n}|DoS_{ij} - Ex_i|$
Calculate hyper entropy: $He_i = \sqrt{S_i^2 - En_i^2}$
Output: Digital characteristics Ex_i, En_i, and He_i of interdisciplinary concept c_i.

3.2.2. Interdisciplinary semantic drift degree calculation based on similarity cloud

Instead of calculating the ISD degree directly, we first calculate the semantic similarity of an interdisciplinary concept between two disciplines based on the similarity cloud. Specifically, we measure the semantic similarity degree of an interdisciplinary concept between two disciplines by the overlap degree of their normal clouds.

Definition 7: Overlapped degree(Wang et al., 2020**).** Assume that there are two clouds c_1 and c_2 in the universe of discourse U. The overlapped degree is defined as Eqs. (7) and (8).

$$ol(c_1, c_2) = \frac{2(\sup(c_{1\alpha}) - \inf(c_{2\alpha}))}{(\sup(c_{1\alpha}) - \inf(c_{1\alpha})) + (\sup(c_{2\alpha}) - \inf(c_{2\alpha}))} \quad if\ c_1 < c_2 \tag{7}$$

$$ol(c_1, c_2) = \frac{2(\sup(C_{2\alpha}) - inf(C_{2\alpha}))}{(\sup(C_{1\alpha}) - \inf(C_{1\alpha})) + (\sup(C_{2\alpha}) - \inf(C_{2\alpha}))} \quad if\ c_1 \geqslant c_2 \tag{8}$$

where $\sup(c_{i\alpha})$ and $\inf(c_{i\alpha})(i = 1, 2)$ are the supremum and infimum of the cloud c_i expectation curve, respectively. Based on the overlapped degree, the similarity between c_1 and c_2 $SIM(c_1, c_2)$ is defined as Eq. (9).

$$SIM(c_1, c_2) = \frac{\mu - \alpha}{1 - \alpha} \times ol(c_1, c_2) \tag{9}$$

where μ represents the certain degree of the intersection of the clouds c_1 and c_2, and α is the certain degree of cloud model "3E" rules.

Definition 8: Degree of interdisciplinary semantic drift (Gulla et al., 2010**).** Assume that there are clouds c_1 and c_2 for an interdisciplinary concept in two disciplines, respectively, the degree of ISD checks the distance between cloud c_1 and c_2. The degree of ISD $(Dr(c_1, c_2))$ is defined as Eq. (10).

$$Dr(c_1, c_2) = 1 - SIM(c_1, c_2) \tag{10}$$

The degree of ISD of concept c between discipline i and discipline j can range from 0 to 1, $0 < Dr(c_i, c_j) < 1$. Otherwise, if $Dr(c_1, c_2) = 0$, it means that the meaning of concept c is the same in disciplines i and j, thus no semantic drift occurs between disciplines i and j. Also, if $Dr(c_i, c_j) = 1$, it means c_i and c_j are homonyms, thus no semantic drift occurs at all, as well.

3.3. Interdisciplinary semantic drift direction identification

New disciplines tend to borrow knowledge from mature disciplines. Plus, the fusion of knowledge from different disciplines usually stimulates innovation, which in turn leads to ISD. The formation process of ISD can be divided into three stages: interdisciplinary knowledge matching, interdisciplinary knowledge flow, and interdisciplinary knowledge fusion, as depicted in Fig. 5.

Interdisciplinary knowledge matching aims to discover relevant knowledge that can be used for new disciplines from mature disciplines. A threshold of the similarity (SIM) of knowledge between the source discipline and the target discipline is used for knowledge selection. After knowledge matching, if the KPE (caused by the uneven distribution of knowledge) of the source discipline is higher than that of the target discipline, knowledge will flow from the source discipline to the target discipline (Hai et al., 2007). The principle behind it is the same as water which will naturally flow from the position with high potential energy to the position with low potential energy. Followed by the interdisciplinary knowledge flow is the knowledge fusion (KF), where the knowledge of the source discipline begins to fuse with the knowledge of the target discipline. Several factors, such as concept deduction, concept

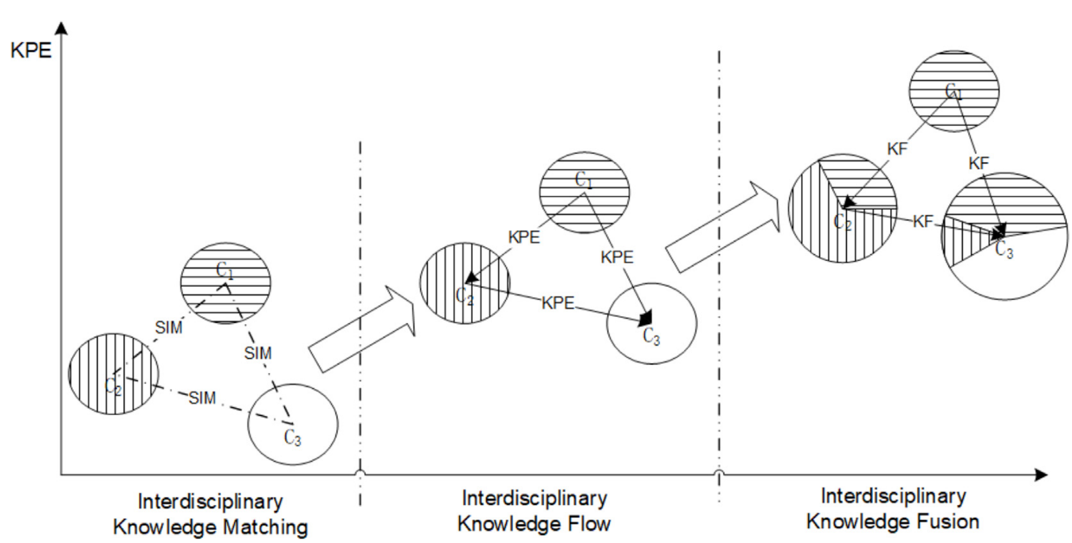

Fig. 5. The formation process of interdisciplinary semantic drift.

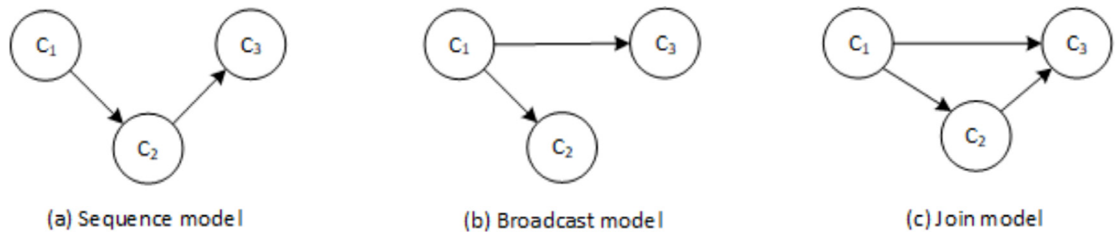

Fig. 6. Models of interdisciplinary semantic drift.

induction, and concept analogy, can affect KF. In other words, the conceptual meaning changes in the interdisciplinary knowledge flow over disciplines lead to the ISD.

From the above analysis, we conclude that, as long as knowledge is relevant between disciplines and there is KPD between disciplines, the concept will drift from the high-KPE discipline to the low-KPE discipline. Therefore, the KPD defines the direction of ISD.

Citation analysis is often used to identify the direction of knowledge flow (Zeng et al., 2021). Additionally, the more citations from the source discipline to the target discipline, the more knowledge flows from the source discipline to the target discipline. Following this idea, we can detect the KPD between disciplines by monitoring cross-disciplinary citations. In this study, the KPD is measured by Eq. (11) based on the difference between the cross-citations of the two disciplines.

Definition 9: Knowledge potential difference (Zhuge et al., 2007**).** Assume that S_i^c and S_j^c are sets of articles containing the concept c in discipline i and j, respectively, the KPD of concept c in source discipline i concerning the receiving discipline $j(KPD(c_i, c_j))$ can be calculated through Eq. (11).

$$\mathrm{KPD}\ (c_i, c_j) = \frac{f_{ij} - f_{ji}}{\max\left(D_i, D_j\right)} \tag{11}$$

where f_{ij} is the number of citations of S_j^c coming from S_i^c, f_{ji} is the number of citations of S_i^c coming from S_j^c, D_i refers to the total number of articles in S_i^c, and D_j refers to the total number of articles in S_j^c.

Unlike potential energy in physics, KPE has no transitive properties. In other words, if $KPD(c_i, c_j) > 0$, and $KPD(c_j, c_k) > 0$, we cannot conclude that $KPD(c_i, c_k) > 0$. This is because a discipline usually obtains knowledge from its nearer disciplines, of which the knowledge is easier to understand and absorb.

3.4. Interdisciplinary semantic drift detection algorithm

Based on the ISD direction identification and the ISD degree calculation approaches presented above, we discuss the algorithm for ISD detection in this section. Generally, there are three different patterns of semantic drift between disciplines: the sequence model, the broadcast model, and the join model as shown in Fig. 6.

- Sequence model. The sequence model is presented in sub-figure (a) in Fig. 6. In this model, the meaning of an interdisciplinary concept c drifts from one discipline to another discipline sequentially. For example, sub-figure (a) shows that the meaning of concept c first drifts from discipline 1 to discipline 2 and then from discipline 2 to discipline 3.
- Broadcast model. The broadcast model is presented in sub-figure (b) in Fig. 6. In this model, the meaning of an interdisciplinary concept c can drift from one source discipline to many other target disciplines directly. For example, sub-figure (b) shows that the meaning of concept c drifts from source discipline 1 to target disciplines 2 and 3.
- Join model. The join model is presented in sub-figure (c) in Fig. 6. In this model, many routes may exist when the meaning of an interdisciplinary concept c drifts from one source discipline to other target disciplines. For example, sub-figure (c) shows that two routes exist between discipline 1 and discipline 3. The first one is that the meaning of concept c drifts from discipline 1 to discipline 2 and then drifts from discipline 2 to discipline 3. The other one is that the meaning of concept c drifts from source discipline 1 to target discipline 3 directly.

We can make the following observations from Fig. 6: for any two disciplines i and j, as long as an interdisciplinary concept c meets the threshold "$KPD(c_i, c_j) > 0$", the ISD occurred between the two disciplines; otherwise, the semantic drift is not happening. The pseudo-code of ISD detection is described in Algorithm 2.

Algorithm 2 Algorithm for detecting interdisciplinary semantic drift.

```
Input: Normal clouds of concept c in different disciplines c_i(Ex_i, En_i, He_i), i ∈ n.
Rank disciplines according to the time when the concept c first occurred in them, and get discipline series data of interdisciplinary
  concept c, that is, C{c_1(Ex_1, En_1, He_1), c_2(Ex_2, En_2, He_2) ... c_i(Ex_i, En_i, He_i) ... c_n(Ex_n, En_n, He_n)}.
for each c_i(Ex_i, En_i, He_i) ∈ C, i ∈ n, do
  for each c_j(Ex_j, En_j, He_j) ∈ C, j ∈ n do
    if i < j <= n, then
      a. calculate Dr(c_i, c_j) using the Eq. (10).
      b. If Dr(c_i, c_j) > 0, Calculate KPD(c_i, c_j) using the Eq. (11).
      e. if KPD(c_i, c_j) > 0, return Dr(c_i, c_j).
    else
     end if
   end for
end for
Output: The degree of ISD from discipline i to j : Dr(c_i, c_j).
```

Journal of King Saud University – Computer and Information Sciences 35 (2023) 101569

4. Experiments and results

4.1. Data collection

In this paper, we use the China National Knowledge Infrastructure (CNKI) as the data source for our experiments, because CNKI is the largest searchable full-text and full-image interdisciplinary journals database in the world.

4.1.1. Dataset for interdisciplinary concept extraction

We first collected all the articles published in two journals of library and information science (LIS) in CNKI: Journal of Library Science in China and Journal of the China Society for Scientific and Technical Information, which are the top two Chinese journals in LIS. We obtained 5,895 papers in total. Then, we extracted keywords from the 5,895 papers. After removing the duplicate keywords, we obtained 11,702 keywords in total. Finally, through keyword searching in CNKI, we obtained the disciplinary distribution of each keyword. After that, interdisciplinary concept extraction experiments were performed based on these keywords and their disciplinary distributions.

4.1.2. Dataset for interdisciplinary semantic drift detection

To test the feasibility of the ISD detection method, we conducted our experiments on the interdisciplinary concept "information entropy". The reasons are as follows: First, "information entropy" has been widely transferred and used in different disciplines through interdisciplinary knowledge communication. Therefore, it can be used as a representative of interdisciplinary concepts in ISD experiments. Second, experts in information science are familiar with the concept of "information entropy", so it is relatively easy for them to evaluate, interpret, and validate our experimental results. Based on the above analysis, we constructed two datasets for direction identification and degree calculation of ISD, respectively.

- To obtain the dataset for identifying the direction of ISD, we first searched for articles on "information entropy" in CNKI. A total of 5,261 articles (dataset D) were retrieved. Then, the articles in dataset D were classified based on the disciplinary classification numbers of papers indexed in CNKI. Therefore, dataset D was further split into several sub-datasets $D(d_1, d_2, \ldots, d_i, \ldots, d_n)$ (d_i are the articles in discipline i). Specifically, the disciplinary distribution of articles on "information entropy" is shown in Table 1. Finally, we identified the cross-reference relationships between disciplines among the 5,261 articles. The above sub-datasets were used to calculate the KPD between disciplines. Based on the KPD, the direction of the ISD can be identified.
- To get the dataset for calculating the degree of ISD, keywords of each discipline were extracted from the keyword lists of the articles in d_i first. Then, removing the duplicate keywords, we obtained a dataset K which includes keywords of each discipline $K(k_1, k_2, \ldots, k_i, \ldots, k_n)$ (k_i are the keywords of discipline i). The dataset K was then used to calculate the degree of ISD.

4.2. Experimental design

We conduct three experiments on the above datasets to test the feasibility and effectiveness of the proposed framework based on NCM for ISD detection. The experiments are described as follows:

- To test the feasibility of the framework, two experiments are performed. (1) Interdisciplinary concept extraction experiment. We first extract interdisciplinary concepts from datasets of LIS based on the proposed interdisciplinary term extraction method. (2) ISD experiment. For the purpose of detecting the ISD, we calculate the degree of ISD based on the similarity cloud and then identify the direction of ISD according to Eq. (11).
- To test the effectiveness of the framework, the third experiment is conducted. (3) Correlation coefficient analysis experiment. We compared the performance of our framework for ISD against three baselines based on correlation coefficient analysis.

Table 1
The disciplines which include the concept "information entropy" in the dataset. Num: the total number of articles in each discipline.

ID	Discipline	Num	ID	Discipline	Num
1	Computer Science	1243	16	Urban Economics	107
2	Control Engineering	705	17	Education	99
3	Quantitative Economics	327	18	Industrial Economics	93
4	Agricultural Economics	269	19	Building Science	79
5	Mathematics	254	20	National Economics	75
6	Electrical Engineering	178	21	Aeronautical Astronautics	73
7	Information and Telecommunication	174	22	Military Science	72
8	General Technology	148	23	Safety	69
9	Business Administration	141	24	Mining Industry	68
10	Mechanical Engineering	136	25	Weapon Engineering	63
11	Communication and Transportation	130	26	Finance	63
12	Physics	123	27	Geology	58
13	Surveying	120	28	Geophysics	58
14	Environment	116	29	Biology	53
15	Hydraulic Engineering	115	30	Science of Library, Information, and Archival	52

Table 2
Scales for evaluating the degree of ISD.

Scale	Indication	Meaning
1	No drift	The connotation and extension of the concept have not changed.
2	Slight drift	The connotation and extension of the interdisciplinary concept have a small extension, which is consistent with the source discipline in general.
3	Moderate drift	The connotation and extension of the interdisciplinary concept have been expanded compared with the source discipline, and new contents have appeared.
4	Heavy drift	The connotation and extension of the interdisciplinary concept in the receiving discipline are very different from the source discipline.

Table 3
Selected interdisciplinary concepts.

ID	Interdisciplinary concept	ID	Interdisciplinary concept
1	Bibliometric analysis	11	Information need
2	Visual analysis	12	Intelligent services
3	Reference services	13	Information resources
4	RDA	14	Information consultation
5	Taxonomy	15	Knowledge management
6	Knowledge graph	16	Personalized services
7	Information resource integration	17	Information fusion
8	Information entropy	18	Open access
9	Resource sharing	19	Knowledge service
10	Data integration	20	Knowledge integration

4.3. Gold standard creation

Since, for our task, there was no appropriate database containing words classified for semantic drift, in this paper, we create the gold standard according to human experts. We first divided the degree of ISD into four levels: no drift, slight drift, moderate drift, and heavy drift, as shown in Table 2. Five experts who are familiar with the field of "information entropy" were invited to evaluate the degree of ISD in different disciplines. To reduce the level of inconsistencies, before the formal evaluation, they were required to pre-evaluate a small amount of data to understand the criteria. We averaged the evaluation results of the five experts as the reference value with which scores given by different ISD analysis methods were compared.

4.4. Baselines

We compare our methods with three concept representation approaches to ascertain the contribution of the NCM to ISD detection. These three approaches can be split into two categories: discrete representation and continuous representation. As for discrete representation approaches, we select VSM to represent an interdisciplinary concept. Each interdisciplinary concept is represented by an N-dimensional vector where N is the number of distinct terms over all the included documents. The second and third approaches are continuous representation methods: Word2Vec and Bidirectional Encoder Representations from Transformers (BERT). Word2-Vec (Mikolov et al., 2013) is a widely used algorithm based on neural networks. It is used to compute vector representation of an interdisciplinary concept in our experiments. BERT (Devlin et al., 2018) is designed to pre-train deep bidirectional representations from the unlabeled text by jointly conditioning on both the left and right context in all layers. BERT has been considered one of the most effective methods for text representation.

4.5. Results and discussion

4.5.1. Interdisciplinary concept extraction

We extract the interdisciplinary concepts by following three steps. We first calculate the values of $Th(k_i)$ and $E(k_i)$ according to Eqs. (1)–(3). Then, keywords, whose $Th(k_i)$ are bigger than 0.23 (γ =0.23) and $E(k_i)$ are bigger than 0.2 (λ =0.2), are selected as interdisciplinary concepts. Lastly, We ask the domain experts to filter out incorrect concepts in the selected interdisciplinary concepts and a total of 516 interdisciplinary concepts are extracted. Table 3 shows 20 of the interdisciplinary concepts in our list. The full list of the interdisciplinary concepts is available on GitHub.

4.5.2. Interdisciplinary semantic drift

After obtaining all the interdisciplinary concepts, the next step is to detect the ISD based on Algorithm 2. Specifically, we first calculate the degree of the ISD using Eq. (10). Then, we calculate the KPD based on the interdisciplinary citations between disciplines using Eq. (11). The greater the KPD, the more knowledge from the source discipline will flow to the target discipline, and the more likely the ISD will happen in the process. Fig. 7 presents the semantic drift process of the concept "information entropy" between different disciplines. In Fig. 7, the direction of the arrow indicates the knowledge flows from the discipline with high KPE to the discipline with low KPE. In addition, the thicker the line, the greater the KPD between disciplines, plus the weight of a link in this network indicates the ISD degree between disciplines and the size of the nodes corresponds to the total number of outgoing links. Therefore, visualizing the ISD chains can comprehensively show experts an outlook of the semantic drift of the target interdisciplinary concept across disciplines.

To identify the maximum and minimum degree of ISD, we plot all the degrees of ISD between Physics and other disciplines in Fig. 8, where the X-axis indicates the discipline IDs, the Y-axis indicates the degrees of ISD, the horizontal dotted line indicates the average of all the degree values, and each of the points indicates the degree of interdisciplinary concept "information entropy" drift from Physics to other disciplines.

Based on the results presented in Fig. 7 and Fig. 8, we make the following conclusions:

- **The longer the disciplinary chain of an interdisciplinary concept transfer, the higher the degree of the ISD.** The reason might be that when an interdisciplinary concept is transferred from one discipline to another, new instances will be accommodated. The accommodated new instances make the extension of the interdisciplinary concept continuously expand, and this leads to a higher degree of ISD. In other words, the introduction of an interdisciplinary concept from one discipline to another leads to changes in its denotation. For example, from Fig. 8 we see that the degree of the interdisciplinary concept "information entropy" drift from Physics to Weapon Engineering is high; the reason is that the interdisciplinary concept "information entropy" was transferred from Physics to Weapon Engineering through a long disciplinary chain (i.e., Physics $\rightarrow$ Mathematics $\rightarrow$Mechanical Engineering $\rightarrow$Weapon Engineering, as shown in Fig. 7). To further analyze this pattern, we visualize the cloud drops of "information entropy" in different disciplines of the disciplinary chain in Fig. 9, which shows that the target discipline inherits the cloud drops of the source disciplines while also integrating new cloud drops of its own discipline. For instance, the cloud drops of the interdisciplinary concept "information entropy" in Weapon Engineering not only inherit from the three source disciplines of Physics, Mathematics, and Mechanical Engineering, but also integrate new cloud drops of its own discipline, as shown in Fig. 9. We also observe that the length of the disciplinary chain is correlated with the diversity of the target interdisciplinary concept's cloud drops: the longer the disciplinary chain of ISD, the more diverse are the target interdisciplinary concept's cloud drops. The diversity of the target interdisciplinary concept's cloud drops can enrich its extension and then lead to a higher degree of ISD.
- **The more source disciplines an interdisciplinary concept comes from, the higher the degree of the ISD.** The reason is that the target discipline concludes the interdisciplinary concept from the source disciplines through interdisciplinary induction. Taking the ISD from Physics to Agricultural Economics as an example, we can see that the degree of interdisciplinary concept "information entropy" drift from Physics to

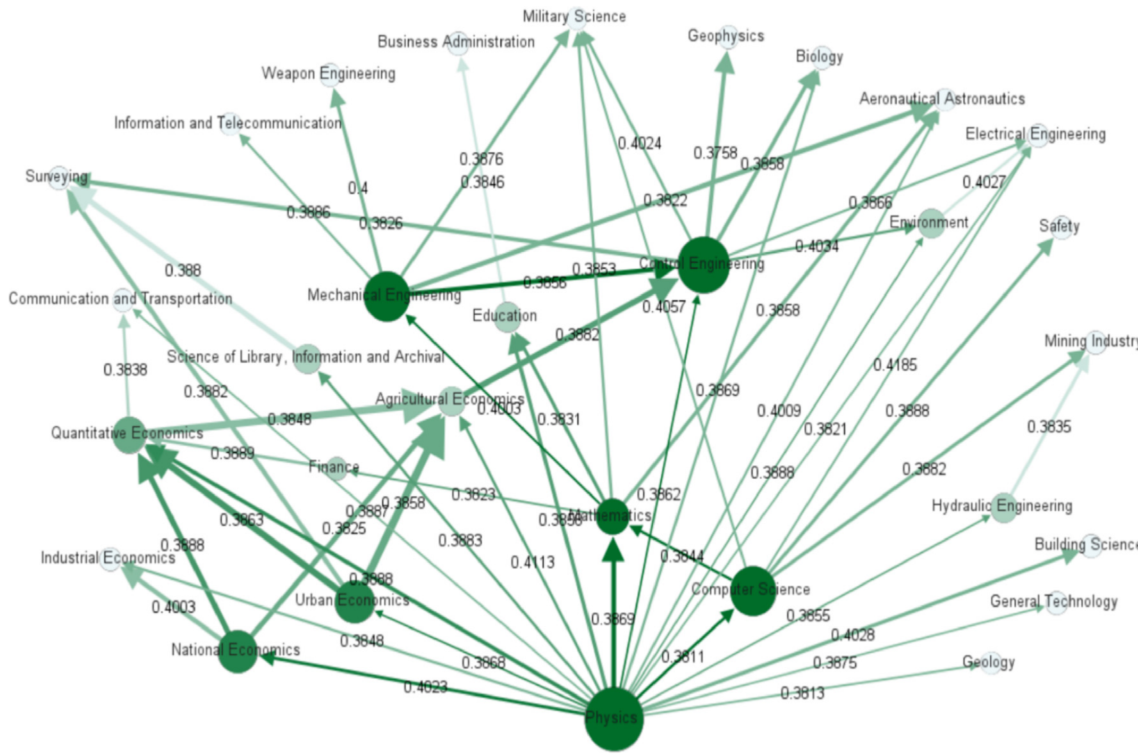

Fig. 7. The ISD of "information entropy" between disciplines.

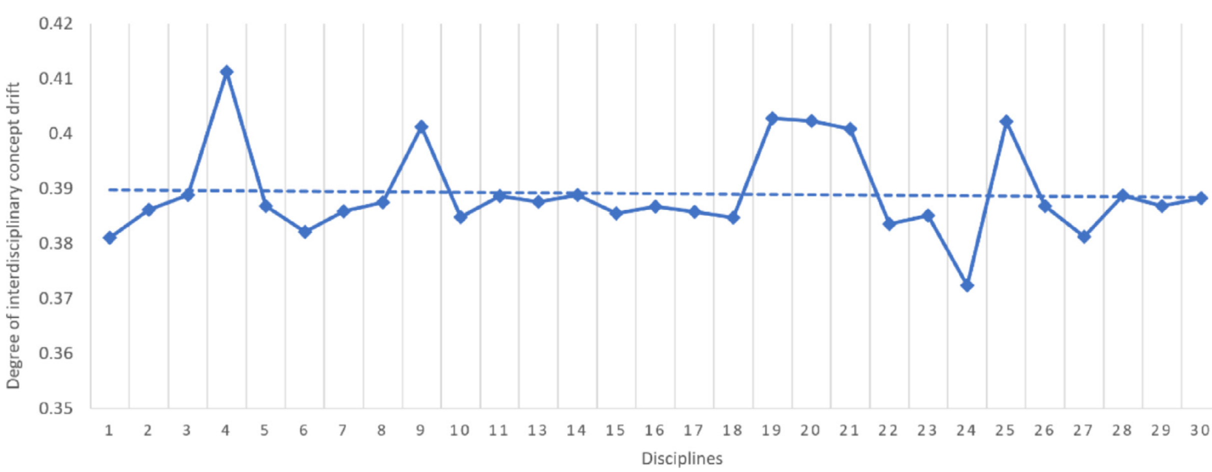

Fig. 8. The degree of interdisciplinary concept "information entropy" drift from Physics to other disciplines.

Agricultural Economics is the highest, as shown in Fig. 8. This is because the interdisciplinary concept "information entropy" is inducted from a lot of source disciplines, including Physics, National Economics, Urban Economics, and Quantitative Economics, as can be seen from Fig. 7. To understand how the structure of the cloud drops affects the degree of ISD, we visualize the cloud drops of "information entropy" in Agricultural Economics in Fig. 10. We find that the more disciplines the target interdisciplinary concept inherits, the greater the heterogeneity of the target interdisciplinary concept's cloud drops, and the higher the degree of ISD. As can be seen from Fig. 10, the cloud drops of "information entropy" in Agricultural Economics inherit the cloud drops from four different disciplines, leading to a higher degree of ISD.

- **The longer the knowledge distance between the source discipline and the target discipline, the higher the degree of ISD.** Intuitively, the longer the knowledge distance between two disciplines, the less overlap or the more difference will exist between the two disciplines. The direct effect of this difference is that the target discipline will inherit less interdisciplinary concept connotation from the source discipline. In other words, the target discipline tends to modify the interdisciplinary concept connotation of the source discipline more. This will lead to a higher degree of ISD. The example of the "information entropy" drift from Physics to National Economics can support this conclusion: the degree of interdisciplinary concept "information entropy" drift from Physics to National Economics is higher than to other disciplines as shown in Fig. 8, since there

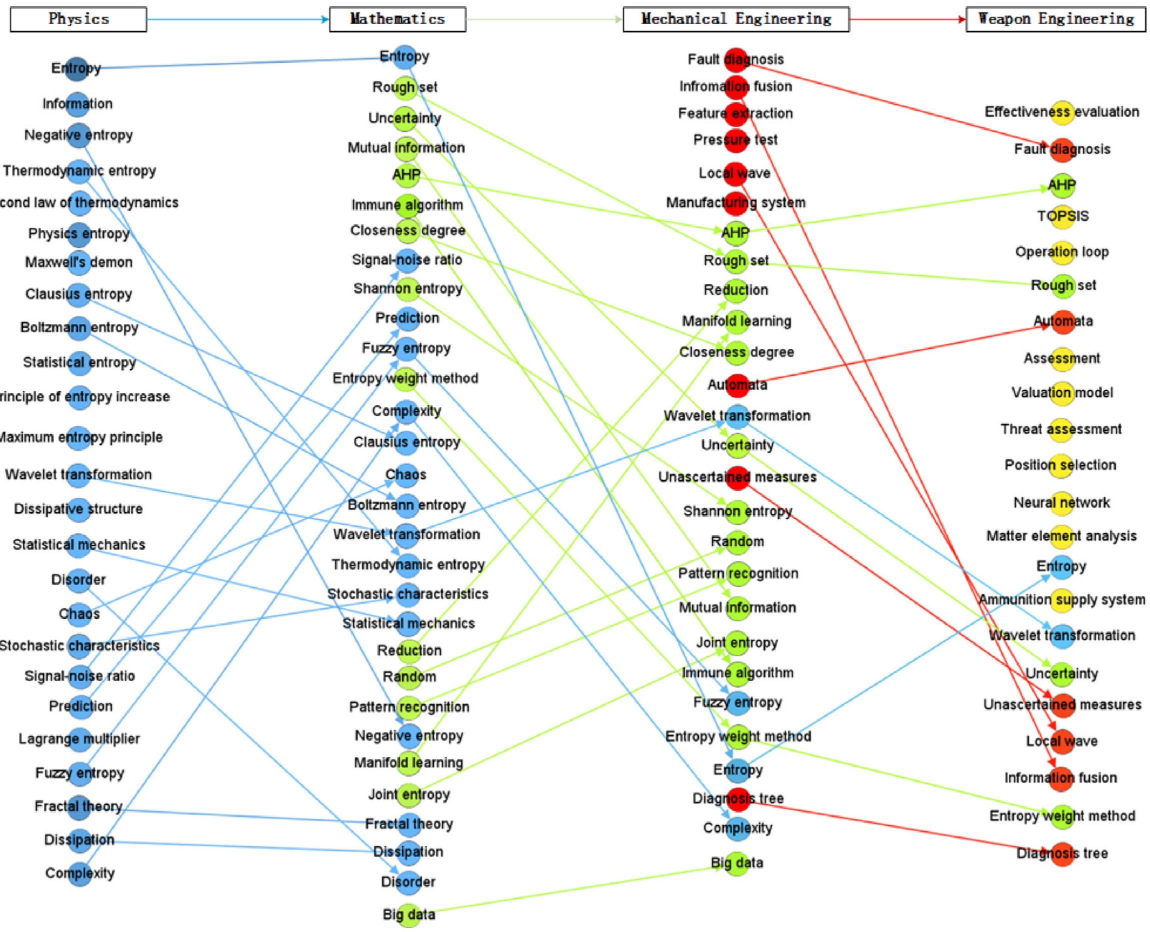

Fig. 9. Cloud drops of "information entropy" in different disciplines of the disciplinary chain.

is a longer knowledge distance between the two disciplines, as shown in Fig. 7. When looking at the cloud drops in this scenario, we observe that the longer the distance between disciplines, the greater the difference of the cloud drops of an interdisciplinary concept in these disciplines. For example, Fig. 11 presents the cloud drops of "information entropy" in National Economics. It shows that the target discipline National Economics only inherits two cloud drops (i.e., Entropy and Boltzmann entropy) from the source discipline Physics.

4.5.3. Correlation coefficient analysis

Table 4 summarizes the ISD degree generated by the proposed framework based on NCM in our paper, the three baselines (VSM, Word2Vec, and BERT), and the experts' judgments. The correlation coefficient analysis between the four algorithms and the experts' judgments is conducted to test their performances.

Pearson's correlation coefficient and Spearman's rank correlation coefficient are widely used in correlation assessment, where Pearson's correlation coefficient is a test statistic measuring the statistical relationship or association between two variables, and Spearman's rank correlation coefficient is a non-parametric statistic used to examine the strength of association between two ranked variables. In the metrics selection, Pearson's correlation coefficient is used for data that fits a normal distribution, and Spearman's rank correlation coefficient is used for data that does not fit a normal distribution. Therefore, in this paper, we first test the normality of the five experimental results of ISD (see Table 4). Considering that each data size of the five experimental results is 30, which is a small sample data, we choose the Shapiro–Wilk test to conduct the normality test, and the results are shown in Table 5. The results in Table 5 indicate that the p-values for the experimental results of VSM, Word2Vec, BERT, and NCM are all less than 0.05, and the p-value for the experimental results of the experts is larger than 0.05, so it is considered that all the experimental results, except the one of the experts, do not obey a normal distribution and, therefore, the Spearman's rank correlation coefficient is chosen as the evaluation indicator. In addition, as the directionality of the experimental results to be tested is known (the experimental result of the experts is larger than the other experimental results), thus a one-tailed Spearman's rank correlation coefficient is used in the correlation assessment of our experimental results.

Based on the above analysis, Spearman single-tailed tests between these four algorithms and the experts' judgments are conducted. The test results are shown in Table 6. From Table 6, we can draw the following conclusions:

- **VSM fails to deal with interdisciplinary concepts' ambiguity and variability**. The Spearman correlation coefficient value between VSM and the experts is 0.294 ($0.2 < 0.294 < 0.39$), with a p-value higher than 0.05 ($0.057 > 0.05$), which indicates that there is no significant correlation between VSM and the experts. Therefore, VSM has a poor performance in ISD detection. The reason is that VSM does not consider the semantic relatedness between concepts.
- **BERT has a better performance than Word2Vec**. The Spearman correlation coefficient value between Word2Vec and the experts is 0.495^{**} ($0.4 < 0.495^{**} < 0.59$), which suggests a moderate correlation. The Spearman correlation coefficient value

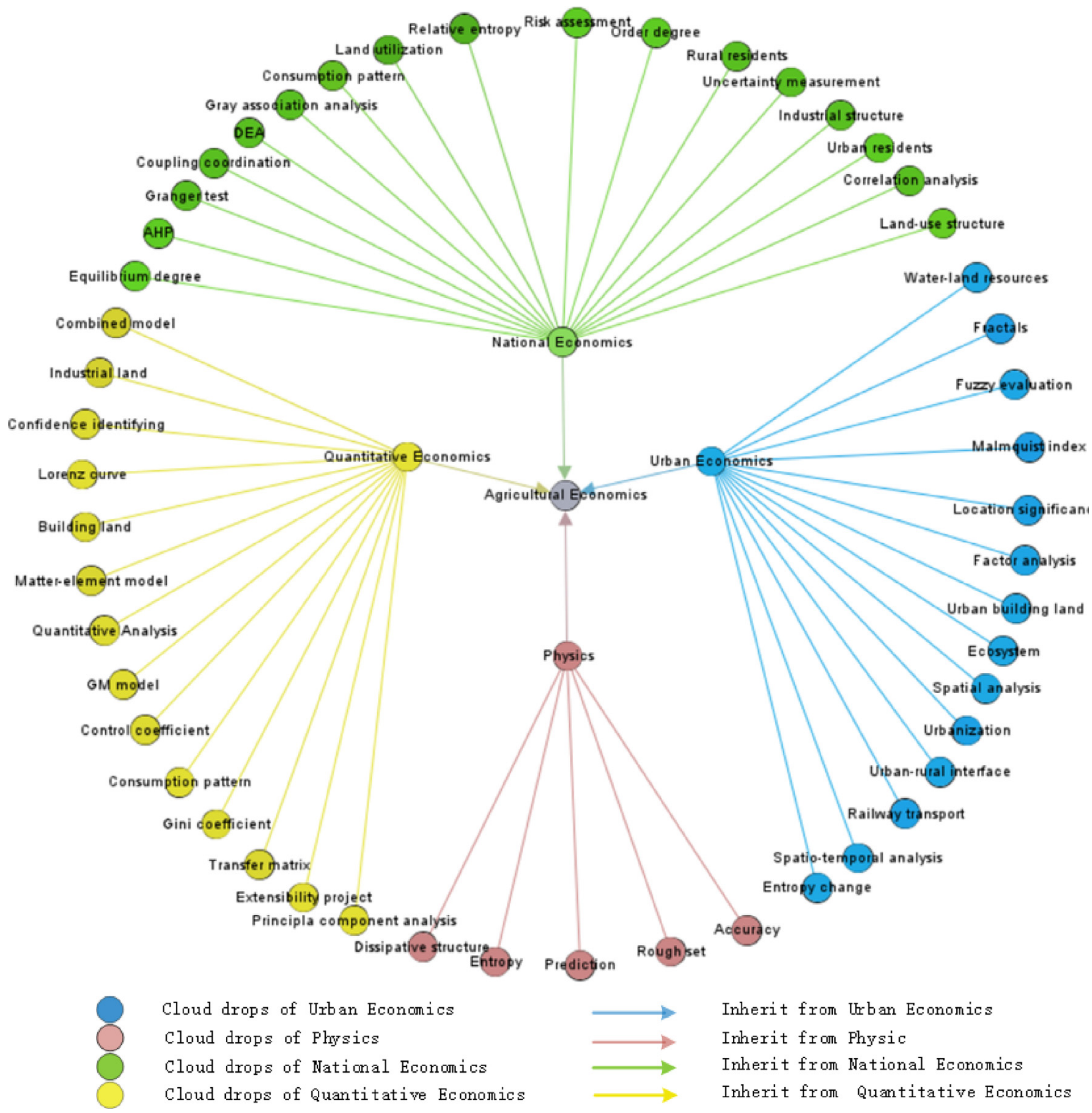

Fig. 10. Cloud drops of "information entropy" in Agricultural Economics.

between BERT and the experts is 0.511**, which is higher than Word2Vec. One reason is that Word2Vec does not address polysemy during concept representation and the other reason is that Word2Vec is also biased toward the term's more frequent sense (Hengchen et al., 2021). Therefore, Word2Vec fails to deal with interdisciplinary concepts' randomness. However, BERT encodes the context of a term and thus can obtain much better results in the concept representation task with the attention mechanism.

- **The performance of our proposed framework based on NCM is better than BERT**. The Spearman correlation coefficient value between the proposed framework based on NCM and the experts lies between 0.80 and 1 ($0.80 < 0.808^{***} < 1$), which is higher than BERT. The results indicate that, compared to BERT, the proposed framework based on NCM has a stronger positive correlation with the experts. The reason is that BERT token representations are highly influenced by target word forms (Laicher et al., 2021). In other words, BERT fails to deal with the uncertainty of a word form for an interdisciplinary concept. Therefore, the proposed framework based on NCM is the only algorithm that achieves comparable results with expert judgments.

Through correlation analysis, we can see that using the proposed framework based on NCM to detect ISD yields a strong correlation with the experts, outperforming approaches based on VSM, Word2Vec, and BERT, which indicates that the ISD detection framework based on NCM proposed in this paper achieves a better result than the above baselines. As interdisciplinary concepts possess uncertain properties, it is difficult to model the problem using precise mathematical models (VSM, Word2Vec, and BERT), which causes errors in ISD detection. Based on this observation, the proposed framework for ISD detection based on NCM can fully employ the advantages of NCM in describing and representing the uncertainty of interdisciplinary concepts. Therefore, it can be used to detect ISD effectively.

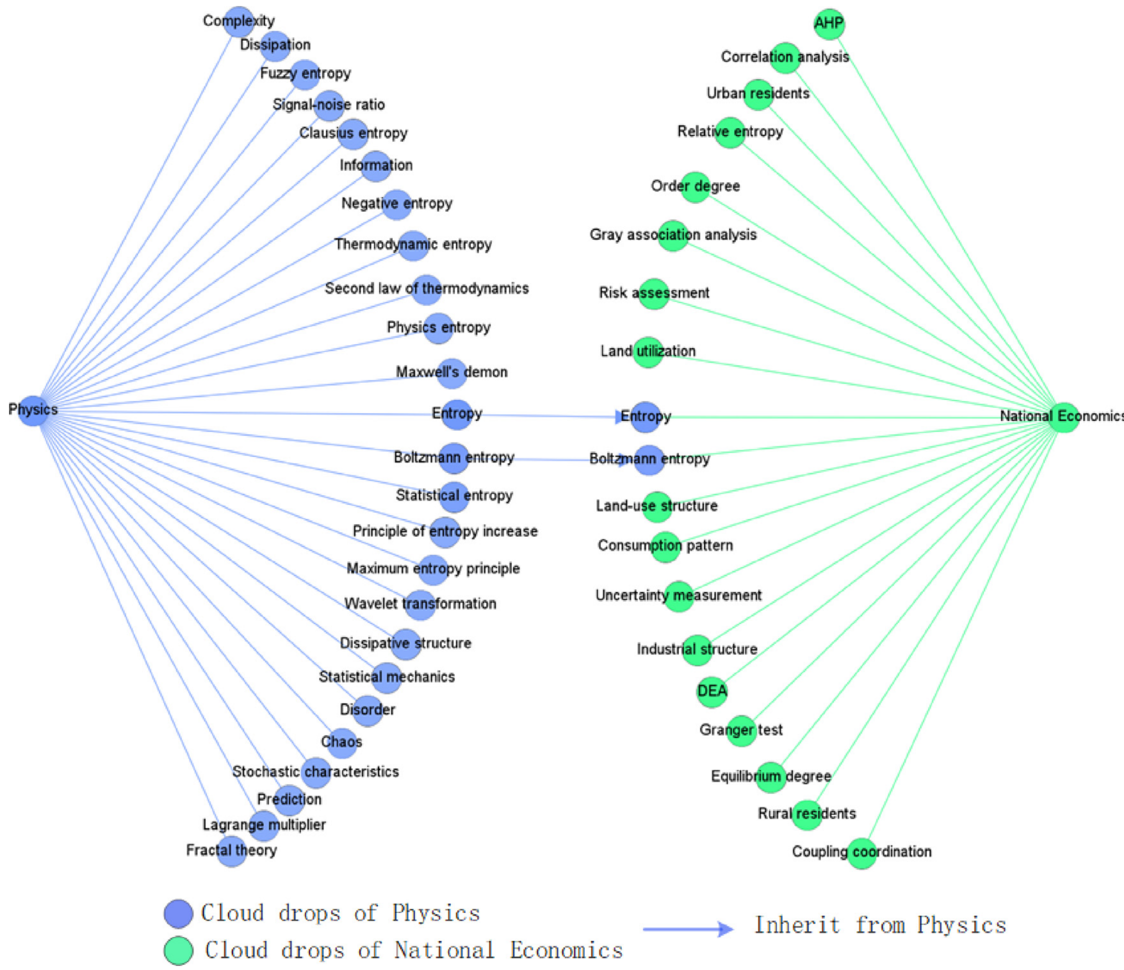

Fig. 11. Cloud drops of "information entropy" in National Economics.

5. Further discussions and implications

The above experimental study has demonstrated the feasibility and effectiveness of the proposed framework for ISD detection in this study. We hereby further discuss the potential implications of the proposed framework in this section.

5.1. Interdisciplinary knowledge organization

KOSs, one of the most crucial applications in the Semantic Web, has drawn increasing attention recently (Greenberg et al., 2021). In almost all types of KOSs, concepts are the central constructs that are used to describe sets of objects with shared characteristics. One of the major problems with interdisciplinary KOSs is the ISDs of their concepts, which may cause serious problems to the interdisciplinary knowledge services that depend on these interdisciplinary KOSs (Szostak et al., 2016; Guimarães, 2017). The ISD detection framework proposed in this paper can track the semantic changes of any interdisciplinary concept over different disciplines, thereby reducing or eliminating the impact of ISD on interdisciplinary KOS-based applications and services.

5.2. Interdisciplinary knowledge retrieval

Interdisciplinary knowledge retrieval (IKR) is essential in interdisciplinary research since it can help researchers obtain relevant knowledge from other disciplines. However, the semantic heterogeneity of concepts between disciplines can significantly reduce the effectiveness of IKR (Bykau et al., 2010). The meaning of a concept might shift when borrowed from one discipline to another (Puustjärvi and Puustjärvi, 2014); such shifting makes the retrieval hardly work across disciplines. Our proposed ISD detection framework can be used to inform the users to modify their queries during the IKR and guide developers to design better applications when ISD is occurring.

5.3. Interdisciplinary knowledge communication

Interdisciplinary knowledge communication combines multiple disciplinary perspectives to produce novel knowledge and problem solutions. However, it requires interdisciplinary knowledge sharing and integration to leverage this potential. Therefore, knowledge in scientific practice is socially constructed meaning. The framework for ISD detection can be used to facilitate researchers gaining a better grasp of essential concepts used by researchers in other disciplines. In other words, it can help researchers involved in interdisciplinary research projects to navigate common grounds regarding the meaning of interdisciplinary concepts and their relation to their respective disciplines.

All in all, studying ISD helps understand how the meaning of a concept shifts from one discipline to another, providing semantic interoperability in interdisciplinary knowledge organization, retrieval, communication, and other relevant applications. Thus, it can enhance interdisciplinary KOSs, improve interdisciplinary knowledge retrieval, overcome the term barrier of interdisciplinary knowledge communication, and accelerate the knowledge innovation process.

Table 4
The ISD degree generated by the framework based on NCM, three baselines, and the experts.

Discipline ID	NCM	VSM	Word2Vec	BERT	Experts
12 → 1	0.3811	0.915	0.776	0.629	1.2
12 → 2	0.3862	0.930	0.837	0.614	1.8
12 → 3	0.3889	0.939	1.022	0.665	3.2
12 → 4	0.4113	0.999	0.907	0.726	3.8
12 → 5	0.3869	0.804	0.412	0.437	1.8
12 → 6	0.3821	0.930	0.526	0.601	1.6
12 → 7	0.3859	0.908	0.539	0.594	2.0
12 → 8	0.3875	0.909	0.731	0.647	2.0
12 →9	0.4013	0.955	1.029	0.673	3.6
12 → 10	0.3849	0.947	0.745	0.601	2.2
12 → 11	0.3887	0.976	0.980	0.748	2.6
12 → 12	0	0	0	0	1
12 → 13	0.3877	0.956	0.831	0.677	2.4
12 → 14	0.3888	0.928	0.874	0.634	2.8
12 → 15	0.3855	0.986	0.982	0.710	2.4
12 → 16	0.3868	0.990	0.738	0.799	3.2
12 → 17	0.3858	0.912	1.030	0.666	3.2
12 → 18	0.3848	0.968	0.809	0.658	3
12 → 19	0.4028	0.923	0.966	0.626	3.4
12 → 20	0.4023	0.974	0.783	0.728	3.6
12 → 21	0.4009	0.954	0.802	0.645	3.2
12 → 22	0.3836	0.968	0.885	0.566	1.6
12 → 23	0.3851	0.979	0.826	0.648	2.4
12 → 24	0.3724	0.980	0.827	0.614	1.2
12 → 25	0.4023	0.963	0.762	0.599	3.2
12 → 26	0.3868	0.965	0.954	0.655	3.2
12 → 27	0.3813	0.971	0.772	0.704	1.6
12 → 28	0.3888	0.950	0.677	0.556	2.6
12 → 29	0.3869	0.882	0.617	0.542	2.4
12 → 30	0.3883	0.804	0.795	0.572	2.6

Table 5
The normality of the five experimental results of ISD analysis regarding the Shapiro—Wilk test.

ISD analysis methods	Significance
VSM	0.000
Word2Vec	0.000
Bert	0.000
NCM	0.000
Experts	0.247

Table 6
Correlation analysis results.

Variables	Significance (one-tailed)	Spearman correlation
VSM-Experts	0.057	0.294
Word2Vec-Experts	0.003	0.495**
BERT-Experts	0.002	0.511**
NCM-Experts	0.000	0.808***

Note: *p <0.05, **p <0.01, ***p <0.001.

6. Summary and future work

With the rapid development of science and technology, interdisciplinary semantic drift occurs more frequently than ever during scientific communication nowadays. This study first investigates the existing efforts for measuring semantic drift. Based on the investigation, we propose a framework for ISD detection based on the normal cloud model. We implement this framework by following four steps: (1) interdisciplinary concept extraction, (2) ISD degree calculation, (3) ISD direction identification, and (4) ISD detection algorithm. To evaluate the feasibility of the proposed framework, we conduct an empirical evaluation using the interdisciplinary concept "information entropy". The evaluation results demonstrate that ISD mainly happens with the interdisciplinary deduction, induction, and metaphor of concepts. We also evaluate the effectiveness of the proposed framework using Spearman single-tailed correlation tests. We believe our proposed framework could benefit multiple applications, as discussed in Section 5.

This study has its limitations, which we hope that future in-depth research studies can address. On the one hand, detecting the direction of ISD is a very complex process, as citation count between disciplines cannot completely represent the direction of semantic drift, since the citation count of an article is influenced by various factors (Onodera and Yoshikane, 2015), including field publication intensity and norms. Therefore, in the future, we will combine other useful information and explore more accurate and effective methods for detecting the direction of ISD. On the other hand, our experiments using the interdisciplinary concepts of LIS leaves open the question as to whether our proposed framework would also hold when detecting other interdisciplinary concepts. Therefore, in the future, we will conduct more case studies on different interdisciplinary concepts and datasets to validate the universality of the proposed framework. Our ultimate goal is to integrate our method into different KOSs to enhance their usefulness and interoperability.

Moreover, the proposed framework will be useful for many downstream applications, such as interdisciplinary knowledge organization, interdisciplinary knowledge retrieval, and interdisciplinary knowledge communication. Therefore, in the future, we also foresee three more extensive research activities. Firstly, based on the ISD detection, we plan to undertake the research of interdisciplinary ontology construction to support interdisciplinary knowledge organization. Secondly, we are going to work on the development of an interdisciplinary knowledge retrieval system to help interdisciplinary researchers access relevant literature from different disciplines. At last, we will commit to interdisciplinary knowledge fusion research to support interdisciplinary knowledge communication among researchers from different disciplines.

Funding

This study is supported by National Social Science Foundation of China (Grant Number 22BTQ102).

CRediT authorship contribution statement

Zhongyi Wang: Conceptualization of this study, Methodology, Software, Writing - Original draft preparation. **Siyuan Peng:** Data curation, Experiments, Writing - Original draft preparation. **Jiangping Chen:** Writing - Original draft preparation, review. **Amoni G. Kapasule:** Writing - Review and editing. **Haihua Chen:** Data curation, Methodology, Writing - Original draft preparation, Project Management.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

de Almeida, M.B., Farinelli, F., 2017. Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. J. Assoc. Informat. Sci. Technol. 68.
Annane, A., Bellahsene, Z., Azouaou, F., Jonquet, C., 2018. Building an effective and efficient background knowledge resource to enhance ontology matching. J. Web Semant. 51, 51–68.
Basile, P., Caputo, A., Semeraro, G., 2014. Analysing word meaning over time by exploiting temporal random indexing. Analysing word meaning over time by exploiting temporal Random Indexing, 38–42.
Basili, R., Moschitti, A., Pazienza, M., Zanzotto, F., 2001. A contrastive approach to term extraction, in: Terminology and Artificial Intelligence Conference (TIA 2001).
Browne, O., O'Reilly, P., Hutchinson, M.C., Krdzavac, N.B., 2019. Distributed data and ontologies: An integrated semantic web architecture enabling more efficient data management. J. Assoc. Informat. Sci. Technol. 70.
Bykau, S., Kiyavitskaya, N., Tsinaraki, C., Velegrakis, Y., 2010. Bridging the gap between heterogeneous and semantically diverse content of different disciplines. In: 2010 Workshops on Database and Expert Systems Applications. IEEE, pp. 305–309.
Capobianco, G., Cavaliere, D., Senatore, S., et al., 2020. Ontodrift: a semantic drift gauge for ontology evolution monitoring. In: CEUR Workshop Proceedings, CEUR-WS. pp. 1–10.
Chen, X., 2014. Exploiting BioPortal as background knowledge in ontology alignment Ph.D. thesis. Miami University.
Cohen, W., Ravikumar, P., Fienberg, S., 2003. A comparison of string metrics for matching names and records. In: Kdd Workshop on Data Cleaning and Object Consolidation, pp. 73–78.
Cross, V.V., Silwal, P., Chen, X., 2013. Semantic similarity measures in ontology alignment. In: 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 442–447.
Darányi, S., Wittek, P., Konstantinidis, K., Papadopoulos, S., Kontopoulos, E., 2016. A physical metaphor to study semantic drift. ArXiv abs/1608.01298.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Eder, J., Wiggisser, K., 2007. Change detection in ontologies using dag comparison. In: International Conference on Advanced Information Systems Engineering. Springer, pp. 21–35.
Feldhoff, B., Stockmann, N., Fanderl, N., Gahle, A.K., Graf, A., Leger, M., Sonnberger, M., 2019. Bridging theories and practices: Boundary objects and constellation analysis as vehicles for interdisciplinary knowledge integration. Sustainability.
Giulianelli, M., Del Tredici, M., Fernández, R., 2020. Analysing lexical semantic change with contextualised word representations. arXiv preprint arXiv:2004.14118.
Greenberg, J., Zhao, X., Monselise, M., Grabus, S., Boone, J., 2021. Knowledge organization systems: A network for ai with helping interdisciplinary vocabulary engineering. Cataloging Classif. Quart. 59, 720–739.
Guimarães, J.A., 2017. Interdisciplinary knowledge organization. J. Assoc. Informat. Sci. Technol. 68, 2287–2289.
Gulla, J.A., Solskinnsbakk, G., Myrseth, P., Haderlein, V., Cerrato, O., 2010. Semantic drift in ontologies. In: WEBIST (2), pp. 13–20.
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.
Hai, Z., Guo, W., Li, X., 2007. The potential energy of knowledge flow. Concurrency Comput.: Practice Exp. 19, 2067–2090.
Hamilton, W.L., Leskovec, J., Jurafsky, D., 2016. Diachronic word embeddings reveal statistical laws of semantic change. ArXiv abs/1605.09096.
Hätty, A., Dorna, M., im Walde, S.S., 2017. Evaluating the reliability and interaction of recursively used feature classes for terminology extraction. In: Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 113–121.
Hazem, A., Bouhandi, M., Boudin, F., Daille, B., 2020. Termeval 2020: Taln-ls2n system for automatic term extraction. In: 6th International Workshop on Computational Terminology (COMPUTERM 2020).
He, T., Fu, W., Xu, J., Zhang, Z., Zhou, J., Yin, Y., Xie, Z., 2022. Discovering interdisciplinary research based on neural networks. Front. Bioeng. Biotechnol. 10.
Hengchen, S., Tahmasebi, N., Schlechtweg, D., Dubossarsky, H., 2021. Challenges for computational lexical semantic change. ArXiv abs/2101.07668.
Hertling, S., Paulheim, H., 2012. Wikimatch: using wikipedia for ontology matching. Ontol. Match. 946.
Hilpert, M., 2008. Germanic Future Constructions: A Usage-based Approach to Language Change, vol. 7. John Benjamins Publishing.
Husein, I.G., Akbar, S., Sitohang, B., Azizah, F.N., 2016. Review of ontology matching with background knowledge. In: 2016 International Conference on Data and Software Engineering (ICoDSE), pp. 1–6.
Khan, S., Safyan, M., 2014. Semantic matching in hierarchical ontologies. J. King Saud Univ.-Comput. Informat. Sci. 26, 247–257.
Kim, Y., Chiu, Y.I., Hanaki, K., Hegde, D., Petrov, S., 2014. Temporal analysis of language through neural language models. ArXiv abs/1405.3515.
Kutuzov, A., Øvrelid, L., Szymanski, T., Velldal, E., 2018. Diachronic word embeddings and semantic shifts: a survey. ArXiv abs/1806.03537.
Laicher, S., Kurtyigit, S., Schlechtweg, D., Kuhn, J., Walde, S.S.i., 2021. Explaining and improving bert performance on lexical semantic change detection. arXiv preprint arXiv:2103.07259.
Lang, C., Wachowiak, L., Heinisch, B., Gromann, D., 2021. Transforming term extraction: Transformer-based approaches to multilingual term extraction across domains. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 3607–3620.
Li, D., Liu, C., Gan, W., 2009. A new cognitive model: Cloud model. Int. J. Intell. Syst. 24, 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 abs/2008.00364.
Ma, Z., Zhang, S., 2020. Risk-based multi-attribute decision-making for normal cloud model considering pre-evaluation information. IEEE Access 8, 153891–153904.
Meroño-Peñuela, A., Hoekstra, R., Guéret, C., Schlobach, S., 2013. Detecting and reporting extensional concept drift in statistical linked data. In: SemStats@ ISWC.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Adv. Neural Informat. Process. Syst. 26, 1–2.
Onodera, N., Yoshikane, F., 2015. Factors affecting citation rates of research articles. J. Assoc. Informat. Sci. Technol. 66, 739–764.
Paharia, N., Pozi, M.S.M., Jatowt, A., 2021. Change-oriented summarization of temporal scholarly document collections by semantic drift analysis. IEEE Access PP, 1–1.
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.
Puustjärvi, J., Puustjärvi, L., 2014. The problem of searching interdisciplinary learning objects. In: Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services, pp. 277–282.
Qin, K., Xu, K., Liu, F., Li, D., 2011. Image segmentation based on histogram analysis utilizing the cloud model. Comput. Mathe. Appl. 62, 2824–2833.
Sagi, E., Kaufmann, S., Clark, B., 2011. Tracing semantic change with latent semantic analysis. Current Methods Historical Semantics 73, 161–183.
Shvaiko, P., Euzenat, J., 2013. Ontology matching: State of the art and future challenges. IEEE Trans. Knowl. Data Eng. 25, 158–176.
Stavropoulos, T.G., Andreadis, S., Kontopoulos, E., Kompatsiaris, I., 2019. Semadrift: A hybrid method and visual tools to measure semantic drift in ontologies. J. Web Semant. 54, 87–106.
Stavropoulos, T.G., Andreadis, S., Kontopoulos, E., Riga, M., Mitzias, P., Kompatsiaris, Y., 2016. Semadrift: A protégé plugin for measuring semantic drift in ontologies. In: Drift-a-LOD@ EKAW, pp. 34–41.
Szostak, R., Gnoli, C., López-Huertas, M., 2016. Interdisciplinary Knowledge Organization. Springer.
Tahmasebi, N., Borin, L., Jatowt, A., 2018. Survey of computational approaches to lexical semantic change. arXiv: Computation and Language.
Takamura, H., Nagata, R., Kawasaki, Y., 2017. Analyzing semantic change in japanese loanwords. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp. 1195–1204.
Thorleuchter, D., Van den Poel, D., 2016. Identification of interdisciplinary ideas. Informat. Process. Manage. 52, 1074–1085.
Wang, P.K., Huang, S., Cai, C., 2021. Dual linguistic term set and its application based on the normal cloud model. IEEE Trans. Fuzzy Syst. 29, 2180–2194.

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. Qual. Reliab. Eng. Int. 36, 1048–1073.
Wardhana, H., Ashari, A., Kartika, A., 2018. Review of ontology evolution process. Int. J. Comput. Appl. 179, 26–33.
Wittek, P., Darányi, S., Kontopoulos, E., Moysiadis, T., Kompatsiaris, Y., 2015. Monitoring term drift based on semantic consistency in an evolving vector field. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.
Wong, W., Liu, W., Bennamoun, M., 2007. Determining termhood for learning domain ontologies using domain prevalence and tendency. In: Proceedings of the Sixth Australasian Conference on Data Mining and Analytics - Volume 70. Australian Computer Society Inc, AUS, pp. 47–54.
Yan, G., Jia, S., Ding, J., Xu, X., Pang, Y., 2019. A time series forecasting based on cloud model similarity measurement. Soft. Comput. 23, 5443–5454.
Yao, J., Wang, G., Wang, L., Zhang, X., Wang, P., 2021. Assessing the spatiotemporal variability of lake water quality using a novel multidimensional shape – position similarity cloud model. J. Hydrol. 599. 126379.
Yao, Z., Sun, Y., Ding, W., Rao, N.S., Xiong, H., 2018. Dynamic word embeddings for evolving semantic discovery. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.
Zeng, B., Lyu, H., Zhao, Z., Li, J., 2021. Exploring the direction and diversity of interdisciplinary knowledge diffusion: A case study of professor zeyuan liu's scientific publications. Scientometrics 126, 6253–6272.
yu Zhang, H., Ji, P., qiang Wang, J., hong Chen, X., 2016. A neutrosophic normal cloud and its application in decision-making. Cognitive Comput. 8, 649–669.
Zhuge, H., Guo, W., Li, X., 2007. The potential energy of knowledge flow. Concurr. Comput.: Practice Exp. 19.