Interdisciplinary Research Classification Based on A Combined Conceptual-Empirical Framework

Shunshun Shi[1,2], Lin Zhang[*1,2,3]

Center for Science, Technology & Education Assessment (CSTEA), Wuhan University, Wuhan, China
Center for Studies of Information Resources, School of Information Management, Wuhan University, Wuhan, China
Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Leuven, Belgium
shishunshun1213@163.com; *linzhang1117@whu.edu.cn

1. Introduction

● The failure of universal indicators of interdisciplinarity

The concept of IDR (Interdisciplinary research) is multi-faced and complex. Various indicators have been presented to measure one or more aspects of this concept. However, recent studies have uncovered that most of the interdisciplinarity indicators are unsatisfying, with deviant results based on supposedly similar indicators.

These studies attest to the pronouncement posted by Ismael Rafols that the universal indicators of interdisciplinarity have failed. In order to confront this failure, a feasible solution with great potential is to construct a theoretical framework about IDR and link it to measurements.

● Classifying interdisciplinarity instead of measuring it directly

Classifying or categorizing is chosen as the guiding idea for the framework construction of IDR, because classifying or categorizing is an effective means to comprehend “beings” in essence, offering resistance against cognitive ambiguity and uncertainty. This study starts from the deconstruction of the concept of IDR, and then propose a classification framework.

● Aims

1. Design an IDR classification framework, spanning from the theoretical to the empirical levels.
2. Combine qualitative and quantitative methods to conduct the classification task.
3. As a typical interdisciplinary field, the COVID-19 research field is selected to prove the validation of our proposed classification framework.

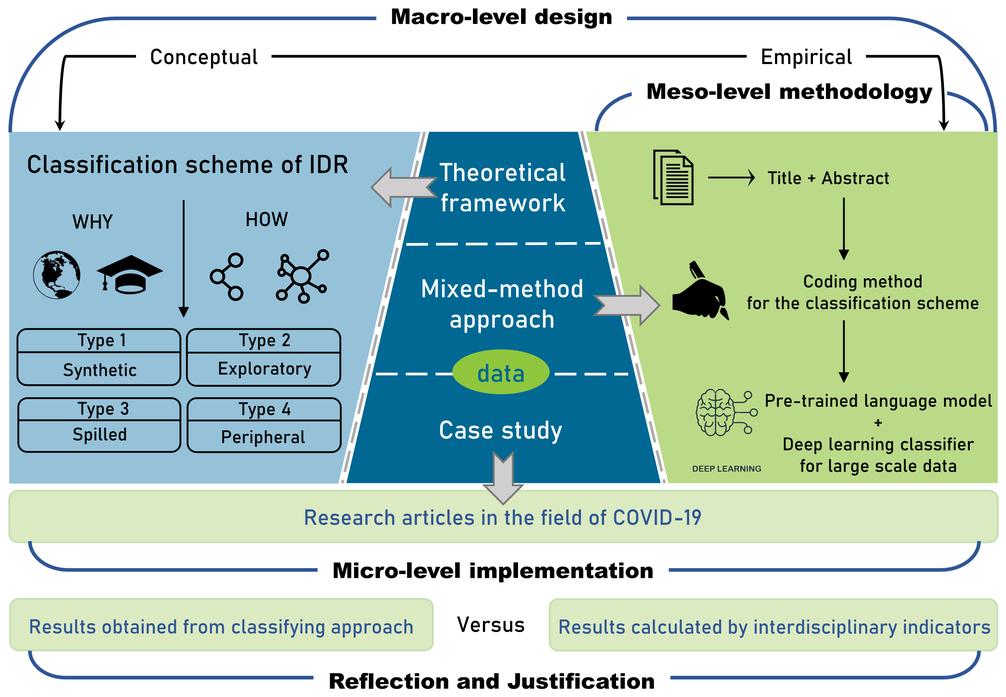

Fig 1. Research design of the proposed IDR classification framework spanning from the theoretical and empirical levels

2. Methods

1 **Coding method** is a systematic process of categorizing, labeling, and organizing qualitative data for analysis. In this study, coding method is used to manually label tags for four types of IDR.

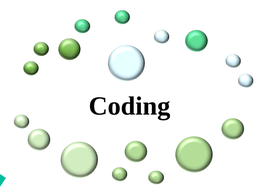

Table 1. Description of four types of IDR based on ‘why’ and ‘how’ dimensions in the context of the target interdisciplinary field

	Type 1 Synthetic	Type 2 Exploratory	Type 3 Spilled	Type 4 Peripheral
Why Dimension	Primary problems	Secondary problems	Subsidiary problems	Virtually irrelevant problems
How Dimension	Integration of knowledge from core disciplines	Integration of knowledge from peripheral disciplines	Integration of knowledge from any disciplines	Integration of knowledge from any disciplines

2 **BERT** is a pre-trained language representation model standing for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.

In this study, a pre-trained BERT model is used to do feature representation work, and the then obtain converted feature vectors for the next classification step.

3 **Convolutional neural network (CNN)** is a form of feed-forward neural network that achieves automatic feature engineering through the optimization of filters. In this study, MLP and CNN are used for the IDR classification task, feeding converted feature vectors obtained from the pre-trained BERT model.

3. Data and Results

3.1 Data description

Titles and abstracts of 4678 articles (COVID-19 related) obtained from the PLOS Library.

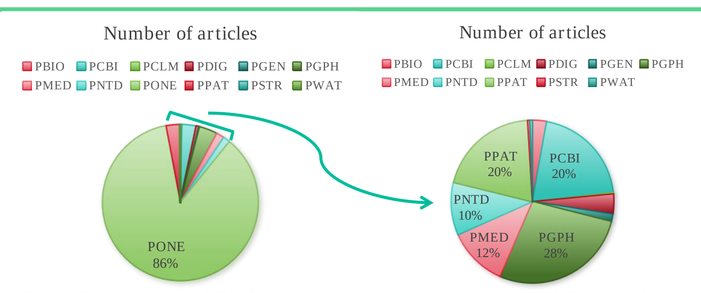

Fig 2. Distribution of PLOS journals by the number of COVID-19 articles

3.2 Coding results

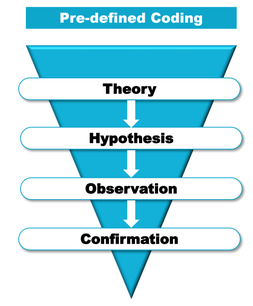

Fig 3. Workflow of the pre-defined coding method

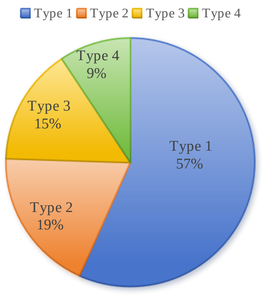

Fig 4. Illustration of article percentage of four types

3.3 Classification results based on deep learning techniques

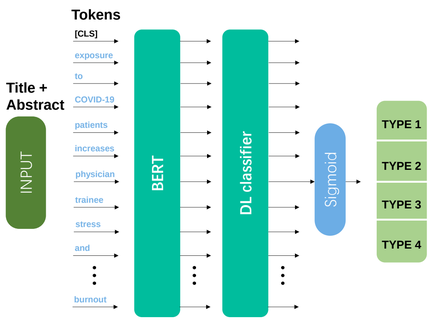

Fig 5. Illustration of the proposed IDR classification model applied in the field of COVID-19

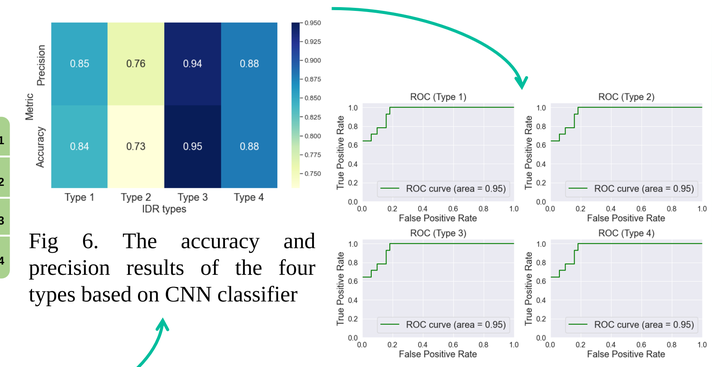

Fig 6. The accuracy and precision results of the four types based on CNN classifier

Fig 7. The ROC results of the four types based on the proposed IDR classification model

4. Conclusions

This study proposes a well-designed classification framework about IDR and link it to a mixed-method approach for implementation to better understand the concept of IDR under its contextualization and thus attempt to overcome the failure of universal interdisciplinarity indicators.

The coding method is used to manually categorize COVID-19 articles based on the '*why*' and '*how*' dimensions, classifying them into one of four types of IDR. Then the labeled articles are utilized to train a deep learning model, comprising pre-trained language model and deep learning classifier to derive the results of the resting unlabeled articles. The final results show that the number of articles of Synthetic type that well-satisfy the definition of IDR accounted for the largest proportion of the total.

5. References

[1] Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, 28(1), 7-22.

[2] Wang, Q., & Schneider, J. W. (2020). Consistency and validity of interdisciplinarity measures. *Quantitative Science Studies*, 1(1), 239-263.

[3] https://www.leidenmadtrics.nl/articles/on-measuring-interdisciplinarity-from-indicators-to-indicating.