Who Should Review Your Proposal? Interdisciplinary Topic Path Detection for Research Proposals

Meng Xiao$^{1,2}$, Ziyue Qiao$^{1,2}$, Yanjie Fu$^3$, Hao Dong$^{1,2}$, Yi Du$^{1,*}$, Pengyang Wang$^4$
Dong Li$^5$, Yuanchun Zhou$^1$
$^1$Computer Network Information Center, Chinese Academy of Sciences, Beijing,
$^2$University of Chinese Academy of Sciences, Beijing,
$^3$Department of Computer Science, University of Central Florida, Orlando, $^4$University of Macau, Macau
$^5$National Natural Science Foundation of China, Information Center, Beijing

$^{1,2}$[shaow, qiaoziyue,donghao]@cnic.cn, $^1$[duyi, zyc]@cnic.cn, $^3$yanjie.fu@ucf.edu, $^4$pywang@um.edu.mo, $^5$lidong@nsfc.gov.cn

ABSTRACT

The peer merit review of research proposals has been the major mechanism to decide grant awards. Nowadays, research proposals have become increasingly interdisciplinary. It has been a longstanding challenge to assign proposals to appropriate reviewers. One of the critical steps in reviewer assignment is to generate accurate interdisciplinary topic labels for proposals. Existing systems mainly collect topic labels manually reported by discipline investigators. However, such human-reported labels can be non-accurate and incomplete. What role can AI play in developing a fair and precise proposal review system? In this evidential study, we collaborate with the National Science Foundation of China to address the task of automated interdisciplinary topic path detection. For this purpose, we develop a deep Hierarchical Interdisciplinary Research Proposal Classification Network (HIRPCN). We first propose a hierarchical transformer to extract the textual semantic information of proposals. We then design an interdisciplinary graph and leverage GNNs to learn representations of each discipline in order to extract interdisciplinary knowledge. After extracting the semantic and interdisciplinary knowledge, we design a level-wise prediction component to fuse the two types of knowledge representations and detect interdisciplinary topic paths for each proposal. We conduct extensive experiments and expert evaluations on three real-world datasets to demonstrate the effectiveness of our proposed model.

1 INTRODUCTION

Currently, research funding is awarded based on proposals’ intellectual, educational, and socio-societal merits. Therefore, scientists submit proposals for their research to open-court competitive programs managed by government agencies (e.g., NSF). Then, proposals are assigned to appropriate reviewers to solicit review comments and ratings. One of the pains of running such a peer-review system is to assign the research proposal to a proper domain, thus allocating a set of domain-related reviewers to advance the review process’s effectiveness and fairness. However, with the increasing number of universities, faculty members, and graduate student hiring, the publications, the ideas, and the submissions of research proposals have been exploding [9, 25]. For example, the Natural Science Foundation of China (NSFC) needs to process more than one million research proposals and find suitable review experts for these proposals every year. Due to this growth momentum, there is an urgent need to bring in AI to assist with proposal categorization, review assignments, panel discussions.

Further, an increasing number of research proposals exhibit an interdisciplinary characteristic, resulting in difficulty in choosing qualified reviewers from one research field. For that, it is critical to classify a proposal into one or more disciplines, or we call it the Interdisciplinary Research Proposal Classification (IRPC) task. Figure 2 shows a toy model of the IRPC task. After analyzing large-scale proposal data from Chinese scientists’ research proposals, we identify three unique data characteristics. These unique properties provide great potential to overcome the challenge in IRPC task:

(P1) The research proposal consists of multi-type of textual data, and different types of text have a distinct meaning. An expert will easily identify the major discipline of a proposal by reading the title, but for a fine-grained category, one has to consider the content in each type. Meanwhile, since the texts in most research proposals vary widely in length and structure, stitching each text together results in a severe loss of information.

(P2) The domain knowledge, peer-reviewers, and the proposal categories are natural to organize into a discipline system. In NSFC, the disciplines category as a taxonomy system, which we call ApplyID$^1$. As shown in Figure 2(3), this system contains thousands of disciplines and sub-disciplines with different levels of granularity, which exhibits a hierarchical structure. Every ApplyID code prefix by a capital letter from A to H, representing eight major

---

$^*$Corresponding author.

---

$^1$htp://www.nsfc.gov.cn/publish/portal0/tab550/
disciplines, followed by zero- to six-digit. Every two-digit number represents a sub-discipline division in a particular granularity. For example, in Figure 2(2), the F refers to the major discipline Information Sciences, the F06 represents Artificial Intelligence, a sub-discipline of Information Sciences, and the F0601 represents Fundamentals of Artificial Intelligence, a sub-discipline of F06.

(P3) Due to the evolution of sciences, the knowledge involved in solving a particular problem has all over the range that a single domain or industry can handle [10]. To evaluate these proposals more fairly, find matching experts based on their related domain, and avoid the destruction of valuable pioneering and interdisciplinary research, it is essential to identify multiple disciplines for those proposals. Figure 2(2) shows a discipline. The example proposal is related to the research field in ApplyID F0601 and B02.

Based on the discussion above, we construct the discipline system and model the task as a top-down iterative process along the discipline hierarchy. For that, we propose Hierarchical Interdisciplinary Research Proposal Classification Network (HIRPCN), a deep learning model with a hierarchical multi-label classification scheme. In HIRPCN architecture, we pursue three unique strategies: First, the model utilizes the historical prediction result in every step to start prediction from any given labels. Second, we use type-token and hierarchical Transformer architecture to preserve type-specific semantic information to handle multiple textual data types. Third, we develop an Interdisciplinary Graph to represent interdisciplinarity and use its topology structure to incorporate the knowledge in each iteration step. In conclusion, our contributions are as follows:

- We identify the critical problems of assigning proposals to its related areas for review. We solve these problems via detecting topic paths on the hierarchical discipline tree.
- We propose HIRPCN, a framework integrating proposals’ semantic information and interdisciplinary knowledge for hierarchical label generation, which can generate one or more topic paths for non-interdisciplinary or interdisciplinary proposals, respectively.
- The experiments and experts’ evaluations show that our model can generate high-quality prediction results and reveal the candidate interdisciplines for research with incomplete labels.
- We online a demo\(^2\) of our model for the critical problem of peer-review assignment on the view of detecting the topic path. Our approach will apply in the funding management in the future and contribute to improving the effectiveness and fairness of the science ecosystem.

2 DEFINITIONS AND PROBLEM STATEMENT

2.1 Important Definitions

2.1.1 A Research Proposal. Applicants write proposals to apply for grants. Figure 2(1) show a proposal includes multiple documents such as Title, Abstract, and Keywords, Research Fields, and more. Let’s denote a proposal by \(A\), the documents in a proposal are denoted by \(D = \{d_i\}_{i=1}^{|T|}\) and the types of each document are denoted by \(T = \{t_i\}_{i=1}^{|T|}\). The \(|T|\) is the total number of the document types, and \(d_i\) is the document of \(i\)-th type \(t_i\). Every document \(d_i\) in the proposal, denoted as \(d_i = [w_i^1, w_i^2, ..., w_i^{|d_i|}]\), is a sequence of words, where \(w_i^k\) is the \(k\)-th word in the document \(d_i\).

2.1.2 Hierarchical Discipline Structure. A hierarchical discipline structure, denoted by \(\gamma\), is a DAG or a tree that is composed of discipline entities and the directed \textit{Belong-to} relation from a discipline to its sub-disciplines. The discipline nodes set \(C = \{C_0 \cup \{C_i\}_{i=1}^H\}\) are organized in \(H\) hierarchical levels, where \(H\) is the depth of hierarchical level, \(C_k = \{c_k^1, c_k^2, ..., c_k^{|c_k|}\}\) is the set of the disciplines in the \(i\)-th level. The \(C_0 = \{\text{root}\}\) is the root level of \(\gamma\). To describe the connection between different disciplines, we introduce \(<\), a partial order representing the \textit{Belong-to} relationship. \(<\) is asymmetric, anti-reflexive and transitive[37]:

- The only one greatest category \textit{root} is the root of the \(\gamma\), 
- \(\forall c_x^x \in C_i, c_y^y \in C_j, c_x^x < c_y^y \rightarrow c_y^y \not< c_x^x\),
- \(\forall c_x^x \in C_i, c_x^x \not< c_x^x\),
- \(\forall c_x^x \in C_i, c_y^y \in C_j, c_x^x \in C_k, c_x^x < c_y^y \wedge c_y^y < c_y^y \not< c_x^x\).

Finally, we define the Hierarchical Discipline Structure \(\gamma\) as a partial order set \(\gamma = (C, \prec)\).

2.1.3 Interdisciplinary Graph. A discipline is a combination of the domain knowledge and topics, and the topic of each proposal is carefully selected [15] and proposed by the scientists and denoted by the Keywords text. By that, we aim to build a Interdisciplinary Graph, denoted as \(G = (C, E)\), to represent the interdisciplinary interactions among disciplines. The \(G\) is a collection of discipline nodes.
C and directed weighted edges $E = \{e_{a\rightarrow b}\}_{d,b=1}^{|C|}$. Each $e_{a\rightarrow b} \geq 0$ represents the interdisciplinarity from discipline $c_a$ to discipline $c_b$.

So, how to measure interdisciplinarity? S. W. Aboelela [1] believe that interdisciplinary research is engaging from two seemingly unrelated fields, which means: (1) these disciplines will have high disparity. (2) the frequency of their related topic will be high. Thus, we adopt the Rao-Stirling [31] (RS) to measure interdisciplinarity. The RS is a nonparametric quantitative heuristic, which widely adopted to measure the interactions and diversity of the discipline system [29–31], ecosystem [3] and energy security [36], which is defined as:

$$RS = \sum_{a,b} (p_a \cdot p_b)\alpha \cdot (d_{a,b})^\beta,$$  

(1)

where the first part $\sum_{a,b} (p_a \cdot p_b)$ are proportional representations of elements $a$ and $b$ in the system (frequency), and the rest $\sum_{a,b} d_{a,b}$ is the distance between the $a$ and $b$ (disparity). The $\alpha$ and $\beta$ are two constant to weighting the two components. We introduce the RS in a micro view to define the weight on $e_{a\rightarrow b}$:

$$e_{a\rightarrow b} = (p_{a\rightarrow b})\alpha \cdot (d_{a\rightarrow b})^\beta,$$  

(2)

where $p_{a\rightarrow b}$ is the proportional of the proposal in discipline $c_a$ that contains same topics in $c_b$, which represents the penetrating strength of $c_a$ to $c_b$. $d_{a\rightarrow b}$ is the topic disparity of $c_a$ in comparing with $c_b$. Their detailed definitions are as:

$$p_{a\rightarrow b} = \frac{\sum a^\alpha (k_i \in (K_a \cap K_b) F_a[k_i])}{\sum a^\alpha F_a[k_i]}, d_{a\rightarrow b} = 1 - \frac{|K_a \cap K_b|}{|K_a|},$$  

(3)

where $\mathbb{1} (\cdot) \rightarrow \{0, 1\}$ is the indicator function, $K_a$ and $K_b$ are Keywords set of $c_a$ and $c_b$, $n$ is the total number of Keywords, and $F_a$ and $F_b$ are the corresponding frequency of the keywords in $K_a$ and $K_b$, appearing in the proposals of their disciplines. In our paper, we assume the two components are equally important and set the $\alpha$ and $\beta$ in Equation 2 as 1.

### 2.2 Problem Formulation

We model the IRPC task in a hierarchical classification schema and use a sequence of discipline-level-specific label sets to represent the proposals’ disciplinary codes, denoted as $L = \{L_0, L_1, L_2, ..., L_{H_{A}}\}$, where $L_0 = \{l_{\text{root}}\}, L_i = \{l_i^j\}_{j=1}^{L_i}$ is the collection of the labels in the label paths on the $i$-th level, i.e., $\forall l_i^j \in L_i \rightarrow l_i^j \in C_1$, $H_{A}$ is the maximum length of label paths. For example, in the Figure 2, the ApplyID codes are F0601 and B02, its labels can be processed as $\{l_{\text{root}}\}, \{F, B\}, \{F06, B02\}, \{F0601\}$. To sum up, given the proposal’s document set $D$ and the interdisciplinary graph $G$, we decompose the prediction process into an top-down fashion from the beginning level to a certain level on the hierarchical discipline structure $\gamma$. Suppose the $k - 1$ ancestors labels in $L$ is $L_{\leq k} = \{L_0, L_1, ..., L_{k-1}\}$, where $L_{\leq 0} = \{L_0\}$, the prediction on level $k$ can be seem as a multi-classification on $L_k$, formed as:

$$\Omega(D, G, \gamma, L_{\leq k}; \Theta) \rightarrow L_k$$

where $\Theta$ is the parameters of model $\Omega$. Also, to define the probability over the length $H_{A}$, we add a particular end-of-prediction label $l_{\text{stop}}$ into the last set in $L$, which enables the iterative process of the model to stop on the proper level when the label $l_{\text{stop}}$ is predicted.

3 INTERDISCIPLINARY LABEL PATH DETECTION

#### 3.1 Overview of the Proposed Framework

Figure 3 show our framework iterate four steps: (1) Semantic Information Extractor (SIE), (2) Interdisciplinary Knowledge Extractor (IKE), (3) Information Fusion (IF), and (4) the Level-wise Prediction (LP). The SIE models the type-specific semantic information (e.g., title, abstract, keywords, research fields) in research proposals. The IKE learns discipline label embedding at current depth by modeling both interdisciplinary graphs and predicted topics labels of shorter depth. The IF fuses the representations of each document in a proposal and topic label embedding. The LP predicts each discipline’s probability to identify the discipline label at the current depth of the three. After completing the iteration, our framework takes predicted discipline labels at the current depth and starts to predict fine-grained discipline next depth.

3.2 Semantic Information Extractor

Figure 4 shows an overview of SIE. The basic purpose is to learn the representations of the textual data of a proposal. A proposal includes a set of documents (e.g., title, abstract, key words, research fields), each of which includes a sequence of words. To learn the representation of a proposal, the SIE is structured to include a Word to Vec layer, a Multiple Positional Encoding (MPE) layer, and a Transformer [33] and 2) a document-level Transformer. Formally, the embedding of a proposal is learned by:

$$D = N_c \times SIEBlock(\{W_i^{(1)}\}_{i=1}^{T}),$$  

(5)

where the matrix $D \in \mathbb{R}^{T \times I \times S}$ is the embedding of the textual data and document type information of the document set of a proposal.
D. \( W_i^{(l)} \in \mathbb{R}^{d_i \times h} \) is the word embedding matrix of document \( d_i \).

The \( N_e \) is the total layer number of SIE Block.

**Our Perspective: Modeling Texts and Document Type for Proposal Embedding.** We consider two essential aspects of a proposal. Firstly, each document (e.g., a title’s texts) in a proposal vary in terms of length, and some document is very short. A short concatenated text will cause a severe loss of long-term dependency. Secondly, different types (e.g., title, abstract, keywords, research fields) of documents in a research proposal have different weights in identifying the disciplines of the proposal. Inspired by the long-text modeling methods [23, 27, 40], we design a hierarchical transformer architecture to model both texts and document type information of a proposal into latent neural embedding.

**Step 1: Word2Vec and Multiple Positional Encoding.** We convert words of each document in a proposal into initial embeddings (denoted by the matrix \( W_i^{(0)} \) of document \( d_i \)) by pre-trained \( h \)-dimensional Word2Vec [24] model and sum with positional encoding to preserve the position information.

**Step 2: Word-level Transformer.** In the first stage of SIEBlock, for each document \( d_i \), we aim to use a Word-level Transformer to extract the contextual semantic information of the document representation matrix \( W_i^{(l-1)} \) of \( d_i \) from previous layer to obtain \( W_i^{(l)} \), formulated as:

\[
W_i^{(l)} = \text{Transformer}_w (W_i^{(l-1)}),
\]

where \( l \) is the current layer number of SIEBlock, the \( W_i^{(l)} \in \mathbb{R}^{d_i \times h} \) is the vanilla document representation matrix and each row is a word embedding.

**Step 3: Document-level Transformer.** In each layer of the Document-level Transformer, we integrate each vanilla document representation matrix with its correlated type information and feed them into the Transformer together:

\[
D_i^{(l)} = \text{Transformer}_d (\{d_i^{(l-1)} \odot W_i^{(l)}\}_{l=1}^{T}),
\]

where \( \text{Transformer}_d (\cdot) \) is the Document-level Transformer, \( d_i^{(l-1)} \) is the type-token vector of \( d_i \) in layer \( (l - 1) \). In the first layer, each type-token vector \( d_i^{(0)} \) will be initialized randomly. Inspired by ViT[8] and TNT[14], we set the \( \odot \) operation as: (1) a Vectorization Operation on \( W_i^{(l)} \) (2) a Fully-connected Layer to transform the vectorized representation from dimension \( d_i \times h \) to dimension \( h \). (3) an Element-wise Add with its type-token vector \( d_i^{(l-1)} \) to get the type-specific representation. \( D_i^{(l)} \in \mathbb{R}^{h} \) is the outputs of current layer, which can be seen as \( |T| \) views of high-level abstraction for proposals. After \( N_e \) times propagation, we can obtain the final output of SIE as \( D = D^{(N_e)} \).

**3.3 Interdisciplinary Knowledge Extractor**

In each iterative prediction step, we hope to use the previously predicted information (or a given partial label) to help the prediction of the next level. Meanwhile, we hope to form the interdisciplinary knowledge within it. Therefore, we propose IKE to learn previous label embeddings. Figure 5 shows an overview of IKE, which mainly consists of multi-layer GCNs and a Readout Layer. From an overall perspective, the label embedding is learned by:

\[
E_{<k} = \text{IKE} (L_{<k}, G),
\]

where the \( E_{<k} \in \mathbb{R}^{k \times h} \) is a label embedding matrix to preserve the interdisciplinary knowledge from previous \( k - 1 \) steps.

**Our Perspective: Modeling interdisciplinary interactions for Discipline Embedding.** It is conceivable to extract the interactions between disciplines on the Interdisciplinary Graph to acquire the interdisciplinary knowledge, just as model people’s behavior by capturing their relation with others on social networks. To utilize the knowledge, we first model the interdisciplinary relation as \( G \), and then use Graph Convolutional Networks (GCNs) to aggregate the neighborhood information into the predicted label embedding.

**Step 1: Grab the Interactions.** Given the sequence of label sets \( L_{<k} \) from previous \( k - 1 \) steps (or from a given partial label from expert), where \( L_{<k} = [L_0, L_1, ..., L_{k-1}] \). Each label set \( L_i \) holds the prediction result in the \( i \)-th level and consists of the discipline labels. We first treat each discipline label in every set as a central node and sample its \( N_g \) hops sub-graph from the interdisciplinary graph, where the \( N_g \) is the layer number of GCNs. Then we feed the sub-graph adjacency weighted matrix and their node features into \( N_g \) layers of GCNs:

\[
H_i = N_g \times \text{GCNL}ayer (E, H_{i}^{(0)}),
\]
where the $H_i^{(0)}$ is a random initialized feature of nodes on the subgraph, which is sampled by the $L_i$ as the central nodes, and the $E_i$ is the weighted adjacency matrix of this sub-graph. The $H_i$ is a hidden output matrix undergoing $N_g$ times GCN layers propagations.

**Step 2: Readout the Features.** Then, we readout the label embeddings of central node set $L_i$ from the output matrix:

$$e_i = \text{Readout}(H_i, L_i),$$

where the Readout layer can be whether a lookup operation or a self-attention layer followed by a mean pooling layer, for simplicity, we set the Readout Layer as the former. The $e_i \in \mathbb{R}^h$ is the final representation of $L_i$. Then we can form the output of IKE as $E_{<k} = [e_0, e_1, ..., e_{k-1}]$.

The $E_{<k}$ can be seen as the representation of the prediction results from the previous $k - 1$ steps integrated with interdisciplinary knowledge. We believe that the knowledge helps the model more inclines to predict the discipline that is highly connected to the historical predicted discipline on the interdisciplinay graph, thereby improving the ability to predict the interdisciplinaries.

### 3.4 Information Fusion

In IF, we use a multi-layer IF Block to integrate the extracted information from SIE and IKE. As the left side of Figure 6 shows, the IF consists of a positional encoding and multiple IF Blocks. We start with its dataflow:

$$S_k = N_d \times IFBlock(E_{<k}, D),$$

where $N_d$ is the total layer number of IF Block, and $S_k \in \mathbb{R}^{k \times h}$ is the fusion matrix that integrated the semantic information and the previous prediction information with interdisciplinary knowledge.

**Our Perspective: Adaptively Utilize Each Part of Information.** There are two strategies we consider in IF. First, the current prediction state should be significantly affected by its previous results. Second, the model should choose the critical part of semantic information adaptively by its current prediction state. To advance those, we divide the IF into two steps.

**Step 1: Positional Encoding and Previous Label Embedding Fusion.** We first sum the $E_{<k}$ with positional encoding to obtain $S_k^{(0)}$ for preserving the order of prediction. Then, in layer-$(l)$ we perform Multi-Head Self-Attention between each prediction result representation to help each element in $S_k^{(l-1)}$ aggregate knowledge from their context:

$$S_k^{(l)} = S_k^{(l-1)} \odot \text{MultiHead}(S_k^{(l-1)} \odot S_k^{(l-1)}, S_k^{(l-1)}),$$

where $\odot$ operation is a Layer Normalization with a Residual Connection Layer. In this step, the MultiHead($\cdot$) adaptively aggregate the context information from every step prediction, and the $\odot$ operation integrate the context knowledge to the label embedding and form the the level-$l$ prediction state as $S_k^{(l)} \in \mathbb{R}^{k \times h}$. With the order information, we believe step 1 of IF can capture the hierarchical dependency and the interdisciplinary knowledge in $L_{<k}$.

**Step 2: Obtain the Semantic Information Adaptively.** As mentioned before, we hope our model can utilize each part of the research proposal adaptively through prediction progress. Thus, we treat the prediction state $S_k^{(l)}$ as Query and the representation of document set $D$ as Key and Value into another Multi-Head Attention to propagate the semantic information:

$$Z^{(l)} = (S_k^{(l)} \odot \text{MultiHead}(S_k^{(l)}, D, D)),$$

$$S_k^{(l)} = Z^{(l)} \odot FC(Z^{(l)}),$$

The MultiHead($\cdot$) adaptively aggregate the semantic information from the research proposal by the current prediction state, and the $\odot$ operation integrate the semantic information to the label embedding and construct the hidden feature $Z^{(l)} \in \mathbb{R}^{k \times h}$ in level-$l$. After a Fully-connected Layer denoted as $FC(\cdot)$ and the $\odot$ operation, we form $S_k^{(l)}$ to hold the fusion information in level-$l$.

After $N_d$ times propagations, we acquire the $S_k = S_k^{(N_d)}$ to represent the output of IF. With the IF, the model will learn the dependency between previous prediction steps and adaptively fuse each part of semantic information by the current prediction progress.

### 3.5 Level-wise Prediction

The right side of Figure 6 demonstrate the prediction in step-$k$. In LP, We feed the fusion feature matrix $S_k$ into a Pooling Layer, a Fully-connected Layer, and a Sigmoid Layer to generate each label’s probability for $k$-th level-wise label prediction. In our paper, we set this Pooling Layer as directly taking the last vector of $S_k$. The formal definition is:

$$\hat{y}_k = \text{Sigmoid}(FC_k(\text{Pooling}(S_k))),$$

where $\hat{y}_k \in \hat{y}_k$ is the predicted probability of the $i$-th discipline $c_i \in C_k$ in level-$k$. The $FC_k(\cdot)$ denotes a level-specific feed-forward network with ReLU activation function to project the input to a $|C_k| + 1$ length vector. After the $\text{Sigmoid}(\cdot)$, the final output $\hat{y}_k$ is the probability of $k$-th level’s labels. Thus, the level-$k$’s objective function $L_k$ can be defined as:

$$L_k(\Theta) = \sum_{i=1}^{|C_i|+1} \left[ \hat{y}_k^i \log(\hat{y}_k^i) + (1 - \hat{y}_k^i) \log(1 - \hat{y}_k^i) \right],$$
where $y^i_k = 1 (t^i_k \in L_k)$, which aims to discriminate whether the $i$-th label $l_i$ belongs to the truth label set or not.

The label prediction start from 1st level with given $L_{<1}$. In the $k$-th level prediction, the label corresponding to the index of the value in the $\tilde{y}_k$ which achieves the threshold will be selected to construct the current prediction result set $L_k$, and the model will append the $L_k$ to $L_{<k}$ to construct $L_{k}$ as current prediction state. If the prediction continues, $L_k$ will form the next previous ancestor $L_{<k}\ldot$. We require that the model can define a distribution over labels of all possible lengths, we add an end-of-prediction label $l_{stop}$ token and set the first element of $\tilde{y}_k$ as the probability of $l_{stop}$ in level-$k$. In the final step, the $l_{stop}$ will include in the prediction result, and the prediction process should end in the level-$(H_k)$, the model will output $L_{H_k}$ as the final result.

4 EXPERIMENTS

In this section, we conduct experiments to evaluate the performance of HIRPCN and answer the following questions: Q1. Is the performance of HIRPCN superior to the existing baseline models? Q2. How is each component of HIRPCN affect the performance? Q3. Can HIRPCN achieve the best performance on the prediction at all levels? Q4. Can the given partial expert advice improve the performance of HIRPCN? Q5. How is the attention mechanism works in each prediction step of HIRPCN? Q6. Can HIRPCN discover the hidden discipline label for interdisciplinary research?

4.1 Experimental Setup

4.1.1 Dataset Description. We collect research proposals written by the scientists from 2020’s NSFC research funding application platform, containing 280683 records with 2494 ApplyID code. 44% are interdisciplinary research. Among those, 16% contain two major discipline labels, and 84% contain one major discipline but show interdisciplinarity in its subdiscipline. The rest research proposals are marked as non-interdisciplinary research by their applicants. We further organized and divided those proposals by their interdisciplinarity into three datasets named RP-all, RP-bi, and RP-differ. RP-all is the overall dataset with all kinds of research proposals. RP-bi consists of the research proposal whether it exhibits interdisciplinarity in the subdiscipline or their major discipline. RP-differ only includes the research proposal with two major disciplines.

4.1.2 Evaluation Metrics. To fairly measure the HIRPCN with other baselines, we evaluate the prediction results with several widely adopted metrics [12, 34, 38] in the domain of Multi-label Classification, i.e., the Micro-F1 (MiF1), and Macro-F1 (MaF1).

4.1.3 Baseline Algorithms. We compare our model with seven Text Classification (TC) methods including TextCNN [4], DPCNN [17], FastText [5, 18], TextRNN [22], TextRNN-Attn [42], TextRCNN [6], Transformer [33] and three state-of-the-art Hierarchical Multi-label Classification (HMC) approaches including HMCN-F, HMCN-R [35], HARNN [16]. We also post three ablation models of HIRPCN as baselines, including w/o Interdisciplinary Graph (w/o IG), w/o Hierarchical Transformers (w/o HT), and w/o All. For all the experiments, we conduct 5-fold cross-validation and report the average recommendation performance.

| Method     | RP-all | RP-bi | RP-differ |
|------------|--------|-------|-----------|
| TextCNN    | 0.456  | 0.168 | 0.426     |
| DPCNN      | 0.376  | 0.108 | 0.364     |
| FastText   | 0.460  | 0.167 | 0.427     |
| TextRNN    | 0.402  | 0.100 | 0.367     |
| TextRNN-Attn | 0.413 | 0.098 | 0.367     |
| TextRCNN   | 0.426  | 0.138 | 0.381     |
| Transformer | 0.570  | 0.083 | 0.339     |
| HMCN-F     | 0.676  | 0.459 | 0.582     |
| HMCN-R     | 0.599  | 0.289 | 0.506     |
| HARNN      | 0.686  | 0.443 | 0.574     |
| w/o ALL    | 0.704  | 0.429 | 0.583     |
| w/o IG     | 0.737  | 0.488 | 0.636     |
| w/o HT     | 0.710  | 0.438 | 0.587     |
| HIRPCN     | 0.748  | 0.512 | 0.648     |

4.1.4 Hyperparameters, Source Code and Reproducibility. We set the SIE layer number $N_s$ to 8, the dimension size $h$ to 64, and the multi-head number to 8. The IKE layer number $N_a$ is set to 1. The IF layer number $N_d$ is set to 8, and the multi-head number is set to 8. We use Word2Vec[24] model with a dimension $(h)$ to 64 to generate the word embedding for each Chinese characters. For the detail of training, we use Adam optimizer[19] with a learning rate of $1 \times 10^{-3}$, and set the mini-batch as 512, adam weight decay as $1 \times 10^{-7}$. The dropout rate is set to 0.2 to prevent overfitting. The warm-up step is set as 1000. We have shared the source code via Dropbox3.

4.1.5 Environmental Settings. All methods are implemented by PyTorch 1.8.1[28]. The experiments are conducted on a CentOS 7.1 server with an AMD EPYC 7742 CPU and 8 NVIDIA A100 GPUs.

4.2 Experimental Results

4.2.1 RJQ1: Overall Comparison. We first evaluate all methods on all datasets in Table 1. In this comparison, we organized every level-wise prediction result flatly, then we used the MiF1 and the MaF1 as the evaluation metrics. From the results, we can observe that:

(1) Our proposed method, HIRPCN achieves the best performance on all datasets in all evaluation metrics, proving that HIRPCN can better solve the challenges mentioned in the IRPC task.

(2) The HMC methods perform better than TC methods. The reason is that the discipline system exhibits a hierarchical structure. HMC methods organize the labels in a hierarchical rather than a flat view, better capturing the hierarchical dependency between labels than TC methods.

(3) It is worth noting that the performance of all the methods decreases on RP-bi and RP-differ compared with that on RP-all. The reason is that interdisciplinary proposals have more labels on each level, making the classification task harder. The performance on RP-bi is generally better than that on RP-differ, indicating that predicting two major disciplines is much more challenging as the disparity on their domain.

3https://www.dropbox.com/sh/x5m1jlcax5p6tk/AAAq-KrGM8eH3uq85BH5LRC0a

Table 1: Experimental results on all dataset. The best results are highlighted in bold. The second-best results are highlighted in underline.
4.2.2 RQ2: Ablation Study. We also discuss the performance of each ablation variant of HIRPCN. The w/o IG removes the IKE component, so in each prediction step, w/o IG will have no awareness of the interdisciplinary knowledge. The w/o HT replaces the SIE component with a vanilla Transformer, which makes the model unable to discriminate the different types of documents and concatenate every word token sequence together. The w/o ALL ablate the IKE and SIE to represent the architecture’s performance. From the Figure 1, we can observe that:

(1) The results of HIRPCN and its three variant models on three datasets prove that both the SIE and IKE components can improve the model’s performance. w/o HT achieves worse results than w/o IG, indicating that SIE contributes more to the task than IKE.

(2) w/o IG performances slightly worse (e.g., -1.5% deterioration on Micro-F1) than HIRPCN on the RP-all dataset, while the margin (e.g., -1.9% & -9.1% deterioration on Micro-F1) becomes larger on RP-bi and RP-differ. This phenomenon proves that the interdisciplin ary interaction topology information plays a significant role in the classification of interdisciplinary proposals.

(3) HIRPCN and w/o IG, both with hierarchical Transformer, are significant superior to other HMC methods on RP-all and RP-bi, showing the advantage of modeling various lengths of the documents and the multiple types of text for proposal embeddings. On RP-differ, w/o IG performs slightly worse than HMCN-F due to the replacement of IKE.

4.2.3 RQ3: Level-wise Performance. We further report the performance of level-wise prediction of HIRPCN and other baselines on the RP-differ dataset. Figure 2 shows the level-wise Micro-F1 and level-wise Macro-F1 of classification results on four levels. From the results, we can observe that HIRPCN outperforms all the baseline methods and the variants on every level, showing the advanced ability on the classification at all granularity. We also observe that the performance of each model tends to decrease with the depth of level increase because the number of categories increases rapidly, resulting in the classification task becoming harder. Lastly, w/o IG demonstrates a relatively slow decreasing tendency compared with other variants, proving the advantage of incorporating interdisciplinary knowledge for fine-grained discipline prediction.

4.2.4 RQ4: Prediction with Given Labels. The Given Labels are a partial label set provided by scientists or funding administrators when they try to fill the ApplyID. As we mentioned in the previous section, we hope the HIRPCN can be an assistant who knows the whole picture of the discipline system. This case study evaluates the improvement when HIRPCN receives the incomplete labels at each level and predicts the rest label. It is also worth noting that other baseline methods use a hidden vector to represent previous prediction results, which cannot initialize the prediction from any given partial label set. Thus, we compared the performance of HIRPCN with different levels of provided expert knowledge on the RP-differ dataset. We evaluate the model by the level-wise MiF1 and the overall MiF1. As you can see in Figure 7, the first row of the figure shows the performance of the HIRPCN on RP-differ when no expert knowledge is given. The rest rows illustrate when the first level, first and second level, first, second and third level are given, how the model improves its performance by using these partial labels. The color of each cell shows the improvement when expert knowledge is given. From the heatmap, we can observe that:

(1) With the incomplete labels given, the overall performance of HIRPCN raised. This phenomenon shows that the architecture guarantees the model can utilize the partial label and can improve the overall performance.

(2) We notice that the given labels can significantly improve the next-level prediction but declines in the remaining level. This phenomenon is reasonable because the model will pay more attention to the previous level than other earlier-level (as the discussion of attention in Appendix A.4 shows).

4.2.5 RQ5: Attention Mechanism Explain. We introduce the attention mechanism in most parts of HIRPCN. In this section, we visualize the attention value learned by the first layer of Word-Level Transformer (further discussion of other attention can be found at Appendix A.4). Figure 8 is a case study of an interdisciplinary research proposal. As we mentioned in Section 3.2, each representation of the word token will be aggregated by different attention values. If a word token gains more SIE’s attention, they will be shading redder in this figure. From the figure, we can observe some interesting phenomena even we take word token as the smallest unit in our model:

(1) As we hope, in the SIE, the model will pay more attention to the critical word but not the stop words, especially in the long text data. For example, in the Abstract, the word “挖掘” (Mining)
and “算法” (Algorithm) will generally gain more attention than the word “是” (Is) and “和” (And). This explains that the HIRPCN can discriminate the importance of each word token incrementally and find the keywords.

(2) Further, the attention value will be high for the word token relevant to the corresponsed discipline of this research proposal. For example, in the Title, the word “肺癌” (Lung Cancer) and “深度挖掘算法” (Deep Mining Algorithm) will gain more attention due to its relation to the Life Sciences and Information Sciences, respectively. We believe this mechanism helps the model to make predictions depending on the semantic information from those critical words. This phenomenon shows that the HIRPCN will extract the useful semantic information from documents by highlight domain-related words.

(3) The self-attention mechanism works well in the text with multiple types and variable-length, whether the “耐药相关” (Drug Resistance) in the Title or the “挖掘” (Mining), “算法” (Algorithm) in the Abstract. We believe this phenomenon proves that the architecture of SIE is superior for handling the complex structure in research proposals.

4.2.6 **RQ6: Hidden Interdiscipline Find.** Among all the wrong cases in RP-all, the HIRPCN classifies part of non-interdisciplinary research to interdisciplinarity, which means except for their labeled discipline, our model predicts an extra discipline. After observing those wrong cases, we noticed that these non-interdisciplinary research proposals were somewhat related to their extra predicted disciplines. To understand this phenomenon, We invited eight fund administrators from NSFC with an insightful understanding and management experience of the eight major disciplines to judge if those extra disciplines can provide helpful information for improving the reviewer assignment. We first group those cases by the extra discipline predicted by HIRPCN. Then we sample 50 research proposals from each group. We send those research proposals to each expert according to the expert’s familiar discipline. If the expert thinks the reviewers from the extra discipline should be considered to supplement to the peer-review group, they will mark it as correct, otherwise wrong. We illustrated the percentage of correct of each extra discipline in Figure 9. From the results, we can observe that:

(1) Overall, experts marked 57% of the samples as correct. Further, all experts comment that the HIRPCN can detect the hidden interdisciplinary research domain reasonably.

(2) Some experts comment that although some cases are marked wrong, most of them explicitly use the keywords from the predicted extra discipline. Experts cannot categorize those cases into interdisciplinary research because their focal point is on the labeled discipline instead of the predicted extra one (e.g., a study on Artificial Intelligence might introduce the idea from Neural Sciences).

(3) Some experts comment that they marked some cases as wrong because the discipline investigator has proposed the labeled discipline to represent an interdiscipline that covered the topic within the predicted discipline. Thus, the labeled discipline is enough to describe the research domain. For example, the discipline Bionics and Artificial Intelligent, denoted by C1005 in Life Sciences, represent the interdiscipline of Bionics and Artificial Intelligent.

We also post four research proposals from different disciplines, which our model predicts a domain from Information Sciences as an extra result. The detail is discussed in Appendix A.6.

5 **RELATED WORKS**

Our work is most related to deep-learning based hierarchical multi-label classification [7] (HMC) methods which leverage standard neural network approaches for multi-label classification problems [11, 13, 35] and then exploit the hierarchy constraint [16] in order to produce coherent predictions and improve performance. Zhang et al. [41] propose a document categorization method with hierarchical structure under weak supervision. The work [32] design a attention-based graph convolution network to category the patent to IPC codes. The work in [21] use the convolution neural network as an encoder and explore the HMC problem in image classification. The works in [26] propose a active learning approach for HMC problem. Aly et al. [2] propose a capsule network based method for HMC problem. Also there are some work focus on co-embedding the classes and entity into vector space for preserve the hierarchy structure. TAXOGAN [39] embedding the network nodes and hierarchical labels together, which focus on taxonomy modeling. In this paper, we inherit the fundamental idea of HMC networks to construct our novel classification methods on the hierarchical discipline system.

6 **CONCLUSION**

This paper proposed HIRPCN, a novel HMC method for the IRPC task on the real-world research proposal dataset. This method extracts the semantic information from each document separately and combines them to achieve a high-level abstraction. Then, the model process each level’s prediction result and neighborhood
topology structure on the pre-defined interdisciplinary graph as interdisciplinary knowledge. The prediction step integrates each part of the research proposal by attention mechanism and generates the next level prediction. The experiments show that our model achieves the best performance on three real-world datasets and can provide the best granular on each level prediction. Other experiments explains the attention mechanism and points out that our model could fix the incomplete interdisciplinary labels under a domain-specific expert evaluation. With the ability to start prediction from any given label, HIRPCN could assist whether researcher or fund administrator to fill the discipline information, which is crucial to improve the reviewer assignation.

REFERENCES

[1] Sally W Aboelela, Elaine Larson, Suzanne Bakken, Olveen Carrasquillo, Allan Formicola, Sherry A Glied, Janet Haas, and Kristine M Gебbie. 2007. Defining interdisciplinary research. Conclusions from a critical review of the literature. Health services research, 42, 1pt1 (2007), 329–346.

[2] Rami Aly, Steffen Remus, and Chris Biemann. 2019. Hierarchical multi-label classification of text with capsule networks. ACL - 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Student Research Workshop (2019), 323–330. https://doi.org/10.18653/v1/p19-2045

[3] Reinette Biggs, Maja Schlüter, Duan Biggs, Erin L Bohensky, Shauna BurnSilver, Georgina Cundill, Vilasdi Dakos, Tim M Dow, Louisa S Evans, Karen Kotschy, et al. 2012. Toward principles for enhancing the resilience of ecosystem services. Annual review of environment and resources 37 (2012), 421–448.

[4] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146. https://doi.org/10.1162/tacl_a_00051 arXiv:1606.04066

[5] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146. https://doi.org/10.1162/tacl_a_00051 arXiv:1606.04066

[6] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146. https://doi.org/10.1162/tacl_a_00051 arXiv:1606.04066

[7] Ricardo Cerri, Rodrigo C Barros, and André CPLF de Carvalho. 2015. Hierarchical classification of gene ontology-based protein functions with neural networks. In 2015 international joint conference on neural networks (IJCNN). IEEE, 1–8.

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).

[9] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staňa Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. 2018. Science of science. Science 359, 6379 (2018).

[10] Jiebo Lu and Pictures of people, dogs, cars, etc. In 2015 international joint conference on neural networks (IJCNN). IEEE, 1–8.

[11] Francesco Gargiulo, Stefano Silvestri, Mario Ciampi, and Giuseppe De Pietro. 2012. Indicators of interdisciplinarity: case studies in bionanoscience. Informetrics 9, 4 (2015), 962–973.

[12] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staňa Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. 2018. Science of science. Science 359, 6379 (2018).

[13] Sally W Aboelela, Elaine Larson, Suzanne Bakken, Olveen Carrasquillo, Allan Formicola, Sherry A Glied, Janet Haas, and Kristine M Gебbie. 2007. Defining interdisciplinary research. Conclusions from a critical review of the literature. Health services research, 42, 1pt1 (2007), 329–346.

[14] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Con...
A APPENDIX

A.1 Graph Convolutional Networks

We adopt Graph Convolutional Networks (GCNs) [20] to extract the network-structured interdisciplinary knowledge from the Co-topic graph. GCN is the most popular and widely-used graph neural networks (GNNs), which is a deep graph information extractor that integrates node features and its local neighbors’ information into low-dimensional representation vectors.

Formally, given a graph $G$, suppose $W \in \mathbb{R}^{N \times N}$ is the adjacency matrix of graph topology, $H^{0} \in \mathbb{R}^{N \times h}$ is the node feature matrix, $N$ is the total number of nodes, and $h$ is the dimension of node feature. GCN uses an efficient layer-wise propagation rule based on a first-order approximation of spectral convolutions on graphs. The $l$-th GCN layer can be formulated as:

$$
H^{(l+1)} = GCN_{layer}(W, H^{(l)}) = \sigma(W^{h}H^{(l)}W^{l})
$$

where $H^{(l+1)} \in \mathbb{R}^{N \times h}$ is the output of the GCN layer, $W^{l} \in \mathbb{R}^{h \times h}$ is the layer-specific parameter matrix, $h$ is the hidden dimension, $\sigma$ denotes an activation function. $W = D^{-\frac{1}{2}}W_{i}D^{-\frac{1}{2}}, W = A + I_{N}$, $I_{N}$ is the identity matrix and $D$ is the diagonal degree matrix of $W$.

In summary, after a $N$-layer GCN’s propagation, formulated as $H^{(N)} = N \times GCN_{layer}(W, H^{(0)})$, we can obtain the final node embedding matrix $H^{(N)} \in \mathbb{R}^{N \times h}$, where each node embedding is integrated with the its $N$-hops neighborhood information.

A.2 Transformer Details

A Transformer [33] model usually has multiple layers. A layer of transformer model (i.e., a Transformer block) consists of a Multi-Head Self-Attention Layer, a Residual Connections and Layer Normalization Layer, a Feed Forward Layer, and a Residual Connections and Normalization Layer, which can be written as:

$$
Z^{(l)} = LN(X^{(l)} + \text{MultiHead}(X^{(l)}, X^{(l)}, X^{(l)})),
$$

$$
X^{(l+1)} = LN(Z^{(l)} + FC(Z^{(l)})),
$$

where $X^{(l)} = [x_{1}^{(l)}, x_{2}^{(l)}, ..., x_{d}^{(l)}]$ is the input sequence of $l$-th layer of Transformer, $s$ is the length of input sequence, $x_{i}^{(l)} \in \mathbb{R}^{h}$ and $h$ is the dimension. $LN(.)$ is layer normalization, $FC(.)$ denotes a two-layer feed-forward network with ReLU activation function, and $\text{MultiHead}(\cdot)$ denotes the multi-head attention layer, which is calculated as follows:

$$
\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_{1}, ..., \text{head}_{d})W^{O},
$$

$$
\text{head}_{i} = \text{Attention}(QW_{i}^{Q}, KW_{i}^{K}, VW_{i}^{V}),
$$

$$
\text{Attention}(Q, K, V) = \text{softmax}(\frac{QKT}{\sqrt{d}})V,
$$

where $W_{i}^{Q}, W_{i}^{K}, W_{i}^{V} \in \mathbb{R}^{h \times h}$ are weight matrices and $d$ is the number of attention heads. After the attention calculation, the $h$ outputs are concatenated and transformed using a output weight matrix $W^{O} \in \mathbb{R}^{dh \times h}$.

In summary, given the input token sequence $X = [x_{1}, x_{2}, ..., x_{s}]$, we first initialize the input matrix $X^{(0)} = [x_{1}^{(0)}, x_{2}^{(0)}, ..., x_{s}^{(0)}]$ before the first layer by a look-up table. Then, after the propagation in Equation 17 and 18 on a $N$-layers Transformer, formalized as $X^{(N)} = N \times \text{Transformer}(X^{(0)})$, we can obtain the final output embeddings of this sequence $X^{(N)} = [x_{1}^{(N)}, x_{2}^{(N)}, ..., x_{s}^{(N)}]$, where each embedding has contained the context information in this sequence. The main hyper-parameters of a Transformer are the number of layers (i.e., Transformer blocks), the number of self-attention heads, and the maximum length of inputs.

A.3 Hyperparameter Selection

We conduct experiments to evaluate the effect of four key hyperparameters, i.e., $N_{g}$, $N_{d}$, $N_{g}$, and the number of self-attention heads in the SIE and IF. We investigate the sensitivity of those parameters in each part of HIRPCN and report the results in Figure 10. From the figure, we can observe that: (1) When the $N_{g}$ and $N_{d}$ varies from 1 to 8, the performance of HIRPCN increases at first then decreases slowly. With more parameters introduced, the model will be likely over-fitting to impact the classification results. (2) We found that the model would have the best performance when IKE sampled a one-hop neighborhood to extract interdisciplinary knowledge. The reason is that the interdisciplinary graph has many edges, and two or more hop sampling strategies will make the sampled sub-graph denser, which will provide unrelated knowledge for IKE. (3) More attention heads will generally improve the performance of HIRPCN. Meanwhile, we find that too many attention heads will make the model perform worse. From the experimental result, we set $N_{g}$ to 8, $N_{d}$ to 8, $N_{g}$ to 1, and the number of self-attention heads to 8 to achieve the best performance.

![Figure 10: Hyper Parameter Selection Study.](image)

A.4 Other Attention Study

Figure 12 shows the attention value of historical prediction results in each step. The Title field is important while predicting the major discipline, and HIRPCN will pay more attention to the Research Field and Abstract to generate the rest labels. Figure 11 shows the attention values to historical prediction results. We notice that the
We believe that, with the HIRPCN, those research proposals can be predicted more accurately. The selected wrong case list in Figure 13. We can explore that each research proposal is highly related to its given major discipline and the average number of labels in each level on three different datasets. Generally speaking, from RP-all to RP-differ, its text data exhibit interdisciplinarity increasingly, and it becomes more challenging to classify correctly in the hierarchical disciplinary structure. In data preprocessing, we filter the incomplete record and remove all the punctuation and padded the length of each text to 200. On the label side, we group the ApplyID codes by their level and add a stop token to the end of each label sequence. As shown in Table 4, when the level goes deeper, since there is more number of disciplines, the classification becomes harder.

### A.5 Data Analysis and Pre-processing

In Table 3, we have counted the lengths of each label sequence and the average number of labels in each level on three different data sets. Generally speaking, from RP-all to RP-differ, its text data exhibit interdisciplinarity increasingly, and it becomes more challenging to classify correctly in the hierarchical disciplinary structure. In data processing, we filter the incomplete record and group the documents in each research proposal by the type of text. We choose four parts of a research proposal as the textual data: Title, Keywords, Abstract, and Research Field. The Abstract part is long-text form, and the average length is 100. The rest three documents are short text. All those documents are critical when experts judge where the proposal belongs. We further removed all the punctuation and padded the length of each text to 200. On the label side, we group the ApplyID codes by their level and add a stop token to the end of each label set sequence. As shown in Table 4, when the level goes deeper, since there is more number of disciplines, the classification becomes harder.

### A.6 Selected Wrong Case

The selected wrong case list in Figure 13. We can explore that each research proposal is highly related to its given major discipline from those cases. Meanwhile, it has strong interdisciplinarity with the Information Sciences, whether the methodology or the idea. We believe that, with the HIRPCN, those research proposals can be reviewed by experts who are knowledgeable in both areas, thus maintaining the fairness of the peer-review system.