7-8-2023

A Dual-view Attention Neural Network for Assigning Industrial Categories to Academic Patents

Zhaobin Liu
City University of Hong Kong, zhaobiliu2-c@my.cityu.edu.hk

Weiwei Deng
South China Normal University, weiweideng@m.scnu.edu.cn

Peihu Zhu
City University of Hong Kong, peihuzhu2-c@my.cityu.edu.hk

Wei DU
Renmin University of China, ahduwei@ruc.edu.cn

Jian Ma
City University of Hong Kong, isjian@cityu.edu.hk

A Dual-view Attention Neural Network for Assigning Industrial Categories to Academic Patents

Completed Research Paper

Zhaobin Liu
College of Business
City University of Hong Kong
Kowloon, Hong Kong
zhaobiliu2-c@my.cityu.edu.hk

Weiwei Deng *
School of Economics and Management
South China Normal University
Guangzhou, China
weiweideng@m.scnu.edu.cn

Peihu Zhu
College of Business
City University of Hong Kong
Kowloon, Hong Kong
peihuzhu2-c@my.cityu.edu.hk

Wei Du
School of Information
Renmin University of China
Beijing, China
ahduwei@ruc.edu.cn

Jian Ma
College of Business
City University of Hong Kong
Kowloon, Hong Kong
isjian@cityu.edu.hk

Abstract

Industrial technology matching events are held by governmental institutions worldwide to promote patent transfer from universities to industries. When collecting academic patents for the matching events, governmental institutions lack professional knowledge for identifying academic patents suitable for various industries. Therefore, previous studies adopted International Patent Classification (IPC) codes assigned by patent examiners to represent patents and mined the industry-related cues through the mapping link between IPC codes and industry categories. However, IPC codes are too general to specifically represent the complex patents, leading to inaccurate tagging. The view of patent inventors (e.g., patent titles and abstracts) contains rich industry-related cues that benefit assigning industrial categories to academic patents. Therefore, we propose a dual-view attention neural network that learns low-dimensional patent representations from the views of patent examiners and inventors and merges the representations for classifying academic patents into suitable industrial categories. Experiments show that the proposed method outperforms benchmark methods.

Keywords: Patent Transfer, Patent Application Analysis, Multi-view Learning

* Corresponding author.

Introduction

Industrial technology matching events are one of the most efficient channels to improve the University-Industry patent transfer and are popular around the world. When preparing for the matching events, governmental institutions need to collect academic patents suitable for various industries. Due to the lack of professional knowledge, governmental institutions spend a lot of time and money analyzing academic patents' applicability to specific industries.

To analyze patents' applicability to specific industries, previous studies adopted IPC codes, the technology codes assigned to patents by patent examiners, to represent patents. Further, the industry-related cues hidden in the IPC codes are mined through the mapping link between IPC codes and industry categories (Dorner & Harhoff, 2018; Evenson & Putnam, 1988; Lybbert & Zolas, 2014; Schmoch et al., 2003; Verspagen et al., 1994). However, as code, IPC is too general to specifically represent the complex patents. For example, the patents with the same IPC codes could have very different functions. Therefore, based on the industry-related cues extracted from IPC codes, academic patents receive the industrial categories with serious bias, leading to a low accuracy of assigning industrial categories to academic patents.

To address the challenge, we propose a new paradigm to analyze patents' applicability to specific industries from a multi-view perspective. Multi views can better describe the complex objects than the single view. Thus, we expect to find another view to the patents as the complement of the patent examiner view, so as to catch more industry-related cues for academic patents. Patent inventors are regarded as the most suitable candidates for evaluating potential applications of their patents (Lybbert & Zolas, 2014). Patent inventors know better about their inventions, such as the research motivation, market requirement, industry product, and other implicit information. Additionally, the above information may be hidden in patent documents, such as patent titles and abstracts. Therefore, in addition to the view of patent examiners, we also consider the view of patent inventors (e.g., patent titles and abstracts). Combining the views of patent examiners and inventors helps better predict the industrial applications of the academic patents. Further, to integrate the two views, we propose a dual-view attention neural network (DANN) to learn the low-dimensional patent representations for potential application analysis. Specifically, the proposed method consists of four modules. The first is the examiner view representation module that defines the examiner view of a patent as the IPC codes assigned to the patent by the examiner. Patent examiners assign IPC codes according to relevant domain knowledge (e.g., IPC architecture knowledge and patent sets corresponding to each IPC) (Righi & Simcoe, 2019; Risch & Krestel, 2019). A large amount of domain knowledge implied in the IPC can help to understand patents with these IPC codes. Therefore, we construct a domain knowledge graph related to the IPC, represent the IPC through knowledge graph embedding, and obtain the examiner view representation of a patent based on the embeddings of its associated IPC codes. The second is the inventor view representation module. Patent titles and abstracts provided by patent inventors contain the main information of patents (Deng & Ma, 2022) and thus can be defined as the inventor view of patents. We obtain the representation of the inventor view by embedding patent titles and abstracts using pre-trained language models. The third is the interaction and fusion module. Patent examiners assign IPC codes to patents after reading the patent documents written by inventors. The relevance between the content and IPC codes of a patent is different. Therefore, we can use the inventor view to refine the examiner view. Specifically, we use the attention mechanism to optimize the weight of different IPC codes in the patent and obtain the refined examiner view representation. Next, we combine the representations of the refined examiner view and the inventor view and input the combined representation to the Deep Neural Network (DNN), which maps the two view representations from two feature spaces to one hidden space. The DNN output is defined as the new patent representation. The fourth is the classification module. We finally classify patents into suitable industrial categories based on the new patent representation.

To evaluate the proposed method, we collect academic patents from the Jiangxi Online Technology Trading Platform where there are many technology matching events. We conduct experiments based on the collected data and experimental results show that the proposed method outperforms baseline methods.

Related Work

This study relates to patent industrial application analysis and multi-view representation learning. Therefore, we review the two types of literature and identify gaps in the literature.

Patent Industrial Application Analysis

To analyze patents' applicability to specific industries, previous studies adopted IPC codes, the technology codes assigned to patents by patent examiners, to represent patents. Further, the industry-related cues hidden in the IPC codes are mined through the mapping link between IPC codes and industry categories. The mapping methods can be clustered into deterministic mapping and probabilistic mapping methods.

Deterministic mapping methods map an IPC code directly to an industrial category. For example, both MERIT (Verspagen et al., 1994) and DG concordances (Schmoch et al., 2003) directly map the IPC into the International standard industrial classification (ISIC). Specifically, MERIT concordance matches IPC subclasses with 22 ISIC classes (Verspagen et al., 1994). The DG concordance assigns 625 IPC subclasses to one of 44 different manufacturing sectors and then links the manufacturing sectors with one or more ISIC codes (Schmoch et al., 2003). Later on, the DG concordance table was updated to link IPC with the current NACE Rev. 2 classification system (Van Looy et al., 2014, 2015). Deterministic mapping methods assign one IPC class to exactly one industrial category, which is rigid and fails to reflect the real situation.

Many studies argued that the technical characteristics represented by an IPC are always related to multiple industrial activities and thus developed probabilistic mapping methods that map one IPC class to several industrial categories with certain probabilities. For example, Yale Technology Concordance established a probabilistic concordance table that maps 8 IPC sectors to 25 industries in the Canadian standard industrial classification system (Evenson & Putnam, 1988). Lybbert and Zolas (2014) designed three weighting schemes when building the mapping relationship between IPC and ISIC: raw, specificity, and hybrid weights. Since the large firms that are active in multiple industries and markets hold the majority of patents, the precision of concordances built on firm-patent linkages data will typically be less than satisfactory. Therefore, Dorner and Harhoff (2018) built the mapping relationship between IPC and "the industry of origin" by using inventor-employee data.

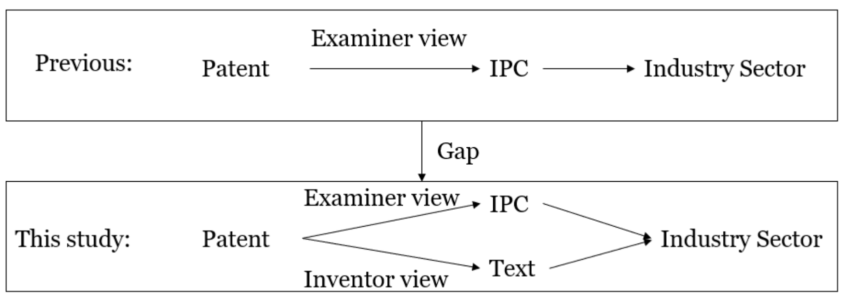

Figure 1. Theoretical Gap of Patent Industrial Application Analysis Literature

According to the mapping link between IPC codes and industrial categories, we can identify a patent's potential industrial application scenario through its IPC codes. However, as code, IPC is too general to specifically represent the complex patents. Therefore, based on the industry-related cues extracted from IPC codes, academic patents receive the industrial categories with serious bias, which leads to inaccurate assignments of industrial categories. To better describe the complex patents, we decide to adopt the multi-view perspective. Thus, we expect to find another view to the patents as the complement of the patent examiner view, so as to catch more industry-related cues for academic patents. Patent inventors are regarded as the most suitable candidates for evaluating potential applications of their patents (Lybbert & Zolas, 2014). Patent inventors know better about their inventions, such as the research motivation, market requirement, industry product, and other implicit information. Additionally, the above information may be hidden in patent documents, such as patent titles and abstracts. Finally, in addition to the view of patent examiners, we also consider the view of patent inventors (e.g., patent titles and abstracts) to analyze the industrial application of patents, as shown in Figure 1.

Multi-view Representation Learning

To fully understand objects, multi-view representation learning has been successfully applied in many tasks, including medical diagnosis prediction, driving behavior analysis, and patent technological field prediction. For example, medical diagnosis prediction can better diagnose disease by analyzing data from the diagnosis and procedure views (He et al., 2020). Analyzing multi-view driving behavior can improve the assessment of drivers (Wang et al., 2018; Wang et al., 2019). The technological field prediction integrated the views of patent inventors, companies, and inventor-company cooperation to represent a patent (Fang et al., 2021).

To the best of our knowledge, there are few attempts to integrate the patent examiner and inventor views to conduct patent application analysis. Because of the complex interaction between different views, simply connecting different views is not conducive to exploring the complementarity between the views. Therefore, it is necessary to develop an advanced method to merge different views for patent industrial application analysis.

Methodology

Given an academic patent **P**, we use $\mathbf{C} = \{\boldsymbol{c_1},\ \boldsymbol{c_2}, \dots \boldsymbol{c_N}\}$ to denote the set of IPC codes of **P**, **D** the combined text of the title and abstract of **P**, $\mathbf{L} = \{\boldsymbol{l_1},\ \boldsymbol{l_2}, \dots \boldsymbol{l_K}\}$ the set of predefined industrial categories. In the patent application analysis task, we predict an industrial category for the academic patent **P** according to **C** and **D** of **P**. Therefore, this is a multi-class classification task. To accomplish the task, we propose DANN to learn the patent representation from both the patent examiner and inventor views for patent industrial application prediction. DANN has four modules, including patent examiner view module, patent inventor view module, interaction and fusion module, and classification module, as shown in Figure 2. The following subsections introduce the four modules in detail.

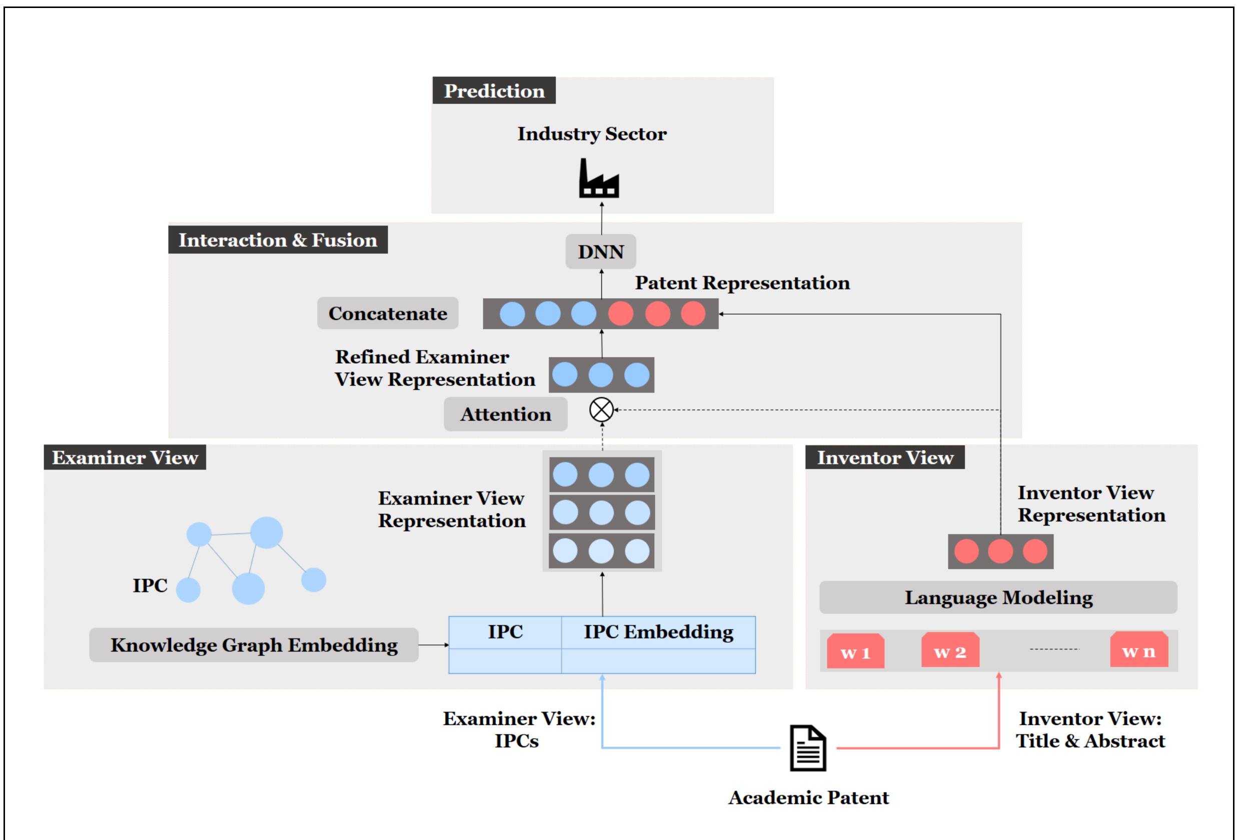

Figure 2. The Framework of DANN

Patent Examiner View Module

We define all IPC codes assigned to a patent as the patent examiner view of the patent. The patent examiner assigns the IPC codes according to the IPC-related domain knowledge (Righi & Simcoe, 2019; Risch & Krestel, 2019). The large amount of domain knowledge implied in the IPC helps to improve the understanding of patents with these IPC codes. Therefore, we extract related domain knowledge to enrich IPC and generate the examiner view representation with the enriched IPC. We adopt knowledge graph to represent IPC-related domain knowledge. Knowledge graph is a network where the nodes represent entities and edges represent relations.

Type	Name	Notes
Entity	*IPC*	The technological field code; The first four digits of the code.
	Patent	The unique number assigned to a patent.
	IKW (IPC KeyWord)	The keyword extracted from the IPC annotation.
	PKW (Patent KeyWord)	The keyword extracted from the patent title and abstract.
Relation	*ClassOf*	*IPC - (ClassOf) ->Patent*
	IKWOf	*IKW-(IKWOf)->IPC*
	PKWOf	*PKW-(PKWOf)->Patent*
	Cite	*Patent-(Cite)->Patent*

Table 1. Descriptions of Entities and Relations in the IPC-related Knowledge Graph

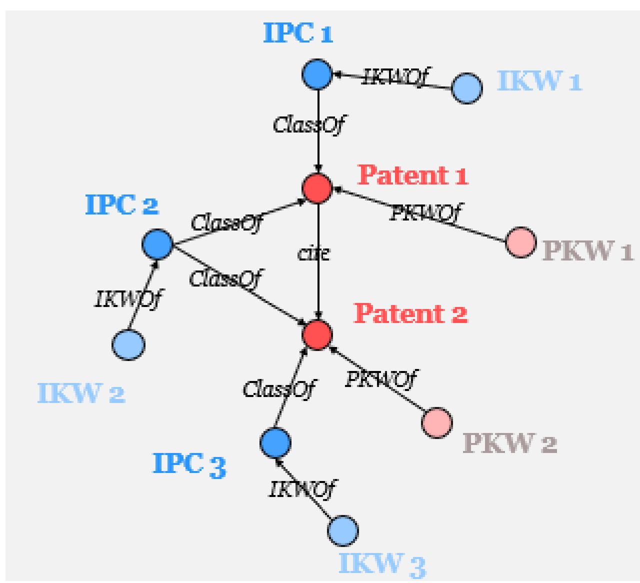

Figure 3. A Toy Example of the IPC-related Domain Knowledge Graph

First, we create a domain knowledge graph to represent the IPC-related domain knowledge, which mainly includes IPC architecture knowledge and patents classified into each IPC. The IPC-related domain knowledge graph consists of four entities and four relations, as shown in Table 1. The entities include *IPC*, *IKW*, *Patent*, and *PKW*. The relations include *IPC - (ClassOf) ->Patent*, *IKW-(IKWOf)->IPC*, *PKW-(PKWOf)->Patent*, and *Patent-(Cite)->Patent*. A patent number represents a unique patent. The first four digits of the IPC code represent the IPC. The rapid automatic keyword extraction (RAKE) algorithm is widely adopted to extract the keywords from patent documents (Du, Jiang, et al., 2021; Du, Wang, et al., 2021; Rose et al., 2012) because this algorithm is unsupervised and domain independent. Therefore, this study employs the RAKE algorithm to extract the keywords from IPC annotation and patent text. Figure 3 shows a toy example of the domain knowledge graph. In Figure 3, there are 3 IPC entities, 2 patent entities, 2 patent keyword entities, 3 IPC keyword entities, 2 *PKWOf* relations, 3 *IKWOf* relations, and 4 *ClassOf* relations.

Second, we use knowledge graph embedding to map all entities and relations of the knowledge graph to a continuous vector space. We employ TransD (Ji et al., 2015) to conduct this mapping job. Specifically, given a fact (h, r, t), TransD introduces three mapping vectors $\boldsymbol{w}_h$, $\boldsymbol{w}_t$, and $\boldsymbol{w}_r$. Two projection matrices are accordingly defined as:

$$\mathbf{M}_r^h = \boldsymbol{w}_r \boldsymbol{w}_h^{\mathrm{T}} + \mathrm{I} \tag{1}$$

$$\mathbf{M}_r^t = \boldsymbol{w}_r \boldsymbol{w}_t^{\mathrm{T}} + \mathrm{I} \tag{2}$$

These two projection matrices are applied on the head entity embedding $\mathbf{h}$ and the tail entity embedding $\mathbf{t}$ respectively to get their projections in the relation embedding space as follows:

$$\mathbf{h}_\perp = \mathbf{M}_r^h \mathbf{h} \tag{3}$$

$$\mathbf{t}_\perp = \mathbf{M}_r^t \mathbf{t} \tag{4}$$

The function of the symbol $\perp$ is to project entity vectors to the vector space of relations. Then, the score function is defined as:

$$f_r(h,t) = -\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2 \tag{5}$$

Figure 4 shows a simple illustration of TransD. The TransD embedding is to enable $\mathbf{h}_\perp + \mathbf{r} \approx \mathbf{t}_\perp$, where $\mathbf{r}$ is the embedding vector of relation r and $\mathbf{h}_\perp$ and $\mathbf{t}_\perp$ denote the embedding vectors of head and tail entities in the relation vector space. We use TransD to embed all entities of the knowledge graph to obtain the matrix $\boldsymbol{E}^{KG}$, which can be regarded as the entity embedding table.

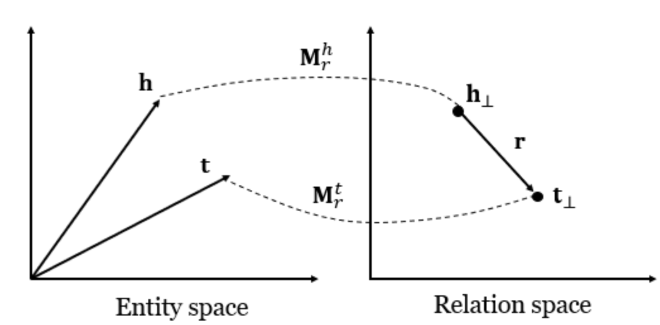

Figure 4. TransD

Finally, according to the set of IPC codes $\mathbf{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \dots \boldsymbol{c}_N\}$ of Patent $\mathbf{P}$, we can get the IPC embedding matrix $\boldsymbol{E}_{\boldsymbol{c}} = [\boldsymbol{E}_{c_1}, \boldsymbol{E}_{c_2}, \dots \boldsymbol{E}_{c_N}]$ of patent $\mathbf{P}$ by looking up the entity embedding table $\boldsymbol{E}^{KG}$. Here, we define $\boldsymbol{E}_{\boldsymbol{c}}$ as patent examiner view representation.

Patent Inventor View Module

we define the title and abstract of a patent as the inventor view of the patent. Bidirectional encoder representations from transformers (Bert) outperforms previous methods in almost all the tasks (Konstantinov et al., 2021). Therefore, we choose the Bert model (Devlin et al., 2018) to embed the **D** of patent **P**.

Specifically, we obtain the outputs of the Bert's first and the last layers. Then, we calculate the avergve value of the [CLS] tokens in these two layers.

$$\boldsymbol{E_D} = \boldsymbol{Bert\_based}\ (\mathbf{D}) \tag{6}$$

Patent inventor view can be represented by the averaged vector.

Interaction and Fusion Module

Patent examiners assign IPC codes to a patent by reading the patent document written by its inventors. Therefore, we can use the inventor view to refine the examiner view, which will help to depict a more accurate patent representation. After that, we integrate these two views.

The examiners often assign more than one IPC code to a patent. The relevance between the given patent text and several IPC codes assigned to the patent differs. Thus, we use the attention mechanism (Vaswani et al., 2017) to optimize the weight of different IPC in the patent. The refined examiner view representation can be calculated as the weighted sum of IPC embeddings. The calculation mechanism is given below.

$$\boldsymbol{q} = \boldsymbol{W^q}\, \boldsymbol{E_D} \tag{7}$$

$$\boldsymbol{K} = \boldsymbol{W^K E_c} \tag{8}$$

$$\boldsymbol{V} = \boldsymbol{W^V E_c} \tag{9}$$

$$\boldsymbol{R_{E_c}} = \boldsymbol{attention}(\boldsymbol{q}, \boldsymbol{K}, \boldsymbol{V}) = \boldsymbol{softmax}\left(\frac{\boldsymbol{q}\,\boldsymbol{K^T}}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{V} \tag{10}$$

Where $\boldsymbol{W^q}$, $\boldsymbol{W^K}$, and $\boldsymbol{W^V}$ are weight matrices and $\boldsymbol{d_k}$ is the dimension of $\boldsymbol{q}$ and $\boldsymbol{k}$.

Because $\boldsymbol{R_{E_c}}$ and $\boldsymbol{E_D}$ are two views from different feature spaces. We need to embed them into a uniform feature space considering cross-domain features. In our model, we use DNN (Hinton & Salakhutdinov, 2006) to extract implicit interaction features. Specifically, we concatenate two representations together as$[\boldsymbol{R_{E_c}}, \boldsymbol{E_D}]$ and pass the concatenated representation to a DNN to generate patent embedding as $\boldsymbol{E_P}$.

Classification Module

Finally, we predict the industrial application scenario for patent **P**.

$$\boldsymbol{P_{IS}} = \boldsymbol{Softmax}\ (\boldsymbol{E_P}) \tag{11}$$

Experiment

Dataset

We collected 17332 academic patents from the Jiangxi Online Technology Trading Platform, where there are many technology matching events hold by Chinese governmental institutions. Currently, the technology matching events in China mainly serve to transfer patents from universities to strategic emerging industries (SEIs) (Prud'homme, 2016). SEIs consist of eight industries, namely, high-end equipment manufacturing (HEEM), digital creative (DC), new materials (NMs), new energy (NE), new energy automobiles (NEVs), biology/biotechnology (BT), new generation IT (NGIT), energy conservation and environmental protection (ECEP) (Kenderdine, 2017). Therefore, we define the above eight industries as the patent industrial application scenarios, which are the labels of the dataset. The technology matching events held in the Jiangxi Online Technology Trading Platform have themes about specific industries. Thus, we assign the industrial labels to academic patents according to the themes of events for which academic patents were selected. For example, we assign BT categories to academic patents selected for the technology matching event with the theme, BT. Table 2 depicts the details of the dataset.

Industrial Category	Number of Patents in each Category	Number of Patents in Training Set	Number of Patents in Test Set
HEEM	2154	1723	431
DC	1697	1358	339
NMs	3303	2642	661
NE	1998	1598	400
NEVs	1428	1142	286
BT	2097	1678	419
NGIT	1865	1492	373
ECEP	2790	2232	558
Total	17332	13866	3466

Table 2. The statistics of the dataset

Baseline Methods

In this section, we will compare the proposed method with baselines for patent application prediction. Here we list the baseline methods in our experiments.

- Deterministic mapping (DM). China National Intellectual Property Administration published a concordance table based on the one-to-one mapping from IPC to SEIs in 2021[1]. By looking up the concordance table, we can assign an industry tag to a patent.
- Bert. Bert (Devlin et al., 2018) is widely used in multi-label classification tasks. When using Bert to classify, we use character-Bert to embed IPC codes, use Chinese-Bert to embed title and abstract, and concate the above two vectors to predict.

Evaluation Metrics

We use precision, recall, and F1 score as the evaluation metrics for this study. We randomly divide the dataset into training and test sets with a ratio of 8:2. We don't apply the cross-validation because of the time-consumption of knowledge graph construction. The three evaluation metrics are mathematically defined as follows:

$$\textbf{Precision} = \frac{TP}{TP+FP} \tag{13}$$

$$\textbf{Recall} = \frac{TP}{TP+FN} \tag{14}$$

$$\textbf{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

Where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

[1] https://www.cnipa.gov.cn/art/2021/2/10/art_75_156716.html

IPC-related Domain Knowledge Graph

We build the knowledge graph using entities extracted from the training set and the IPC annotations from the official website. Table 3 describes the statistics of the IPC-related domain knowledge graph.

Type	Name	Total Number
Entities	*Patent*	13865
	IPC (4 digits/subclass)	485
	IKW	1243
	PKW	15657
Relations	*ClassOf*	26856
	IKWOf	2585
	PKWOf	47454
	Cite	583

Table 3. Statistics of the IPC-related domain knowledge graph

Experimental Results

DANN was implemented in PyTorch. The number of epochs, learning rate, and dropout rate were set to 10, 0.00002, and 0.3, respectively. We compare our method against the baseline methods. Table 4 presents the experimental results of different methods. We have the following findings. First, the DM method performs the worst, which indicates that compared with the dual-view based method, the single-view based method has less capability to represent a patent. Second, the proposed method outperforms all baseline methods in terms of precision, recall, and F1 score. The results provide strong evidence that our proposed method is effective and robust for patent industrial application prediction. The detailed results of the proposed method have been shown in Table 5. In summary, the precision, recall, and F1 score of most industrial categories are about 80%. This means that the proposed method is reliable in classifying academic patents into suitable industrial categories.

Method	Precision	Recall	F1 score
DM	0.6283	0.6191	0.6237
Bert	0.7393	0.7263	0.7327
DANN	**0.7871**	**0.7849**	**0.7856**

Table 4. Comparison between Different Classification Methods

Industry	Precision	Recall	F1 score
HEEM	0.6876	0.7537	0.7192
DC	**0.8424**	0.8111	**0.8265**
NMs	0.7908	0.7718	0.7812
NE	0.8321	0.8198	0.8259
NEVs	0.7762	0.7439	0.7597
BT	0.7905	**0.8375**	0.8133
NGIT	0.7587	0.7311	0.7447
ECEP	0.8183	0.8099	0.8141
Macro Avg	0.7871	0.7849	0.7856

Table 5. The Detailed Results of the Proposed Method

Ablation Analysis

To examine the performance gains obtained by DANN from its three modules, including patent examiner view, patent inventor view, and attention-based interaction and fusion, we conduct an ablation study to compare DANN with three variants, such as DANN-examiner, DANN-inventor, and DANN-avg. Specifically, DANN-avg is to use average strategy to fusion several IPC embeddings to obtain the refined examiner view representation, without considering attention-based interaction and fusion between the examiner view and the inventor view representations. As shown in table 6, we have several findings as follows. First, DANN-avg and DANN outperform DANN-examiner and DANN-inventor, which indicates that the combination of the examiner and inventor views can provide more industry-related cues than the single examiner view or the single inventor view. Second, DANN has better performance than DANN-avg, which infers that interaction and fusion of two views is essential.

Method	Precision	Recall	F1 score
DANN-examiner	0.7034	0.7012	0.7023
DANN-inventor	0.7235	0.7147	0.7191
DANN-avg	0.7543	0.7485	0.7514
DANN	**0.7871**	**0.7849**	**0.7856**

Table 6. Comparison between Variant DANN Methods

Conclusion

Industrial technology matching events are popular worldwide because they are efficient to improve University-Industry patent transfer. However, the event organizers failed to effectively identify the academic patents needed due to the lack of professional knowledge. Therefore, we propose to analyze the patents from the patent examiner and inventor views to improve the efficiency and accuracy of patent industrial application analysis. Further, we propose a novel deep learning method. Experimental results based on a real-world dataset demonstrate that the proposed method is effective and robust.

The theoretical implications of this study are summarized. First, we develop a new paradigm for patent industrial application analysis from a multi-view perspective. We learn patent representations from the patent examiner and inventor views. Second, considering the interaction between the two views, we propose a new method called Dual-view Attention Neural Network. This method has four modules. The first is the examiner view representation module. We first build the IPC-related domain knowledge graph and use TransD to represent the IPC codes. After that, we obtain the examiner view representation composed of several IPC embeddings according to the IPC codes assigned to a patent. The second is the inventor view representation module. We use Bert to embed the title and abstract of the patent. The third is the interaction and fusion module. We use the attention mechanism to optimize the weight of different IPCs in the patent and get the refined examiner view representation. Further, we combine the representations of the refined examiner and inventor views and input the combined representation to the DNN. The fourth is the classification module. We finally predict the potential industrial application scenario based on the DNN output, the new patent representation.

The managerial implication of this study is to assist governments and public research organizations to analyze patent industrial applications by improving the efficiency and accuracy of patent analysis.

Acknowledgements

This research was supported by grants from the Guangdong Basic and Applied Basic Research Foundation (No. 2022A1515110677), the Natural Science Foundation of Guangdong Province (No. 2022A1515011363), and the Shenzhen Commission of Science and Technology (No. 9240067).

References

Deng, W., & Ma, J. (2022). A knowledge graph approach for recommending patents to companies. *Electronic Commerce Research*, *22*(4), 1435-1466.

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*.

Dorner, M., & Harhoff, D. (2018). A novel technology-industry concordance table based on linked inventor-establishment data. *Research Policy*, *47*(4), 768-781.

Du, W., Jiang, G., Xu, W., & Ma, J. (2021). Sequential patent trading recommendation using knowledge-aware attentional bidirectional long short-term memory network (KBiLSTM). *Journal of Information Science*, 01655515211023937.

Du, W., Wang, Y., Xu, W., & Ma, J. (2021). A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis. *Scientometrics*, *126*, 9369-9391.

Evenson, R., & Putnam, J. (1988). The Yale-Canada patent flow concordance. *D, Yale University*.

Fang, L., Zhang, L., Wu, H., Xu, T., Zhou, D., & Chen, E. (2021). Patent2Vec: Multi-view representation learning on patent-graphs for patent classification. *World Wide Web*, *24*(5), 1791-1812.

He, Y., Wang, C., Li, N., & Zeng, Z. (2020). Attention and memory-augmented networks for dual-view sequential learning. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 125-134).

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504-507.

Ji, G., He, S., Xu, L., Liu, K., & Zhao, J. (2015). Knowledge graph embedding via dynamic mapping matrix. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (pp. 687-696).

Kenderdine, T. (2017). China's industrial policy, strategic emerging industries and space law. *Asia & the Pacific Policy Studies*, *4*(2), 325-342.

Konstantinov, A., Moshkin, V., & Yarushkina, N. (2021). Approach to the Use of Language Models BERT and Word2vec in Sentiment Analysis of Social Network Texts. *Studies in Systems, Decision and Control*, *337*, 462-473.

Lybbert, T. J., & Zolas, N. J. (2014). Getting patents and economic data to speak to each other: An 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, *43*(3), 530-542.

Prud'homme, D. (2016). Dynamics of China's provincial-level specialization in strategic emerging industries. *Research Policy*, *45*(8), 1586-1603.

Righi, C., & Simcoe, T. (2019). Patent examiner specialization. *Research Policy*, *48*(1), 137-148.

Risch, J., & Krestel, R. (2019). Domain-specific word embeddings for patent classification. *Data Technologies and Applications*, *53*(1), 108-122.

Rose, S. J., Cowley, W. E., Crow, V. L., & Cramer, N. O. (2012). Rapid automatic keyword extraction for information retrieval and analysis. In: US Patent 8,131,735.

Schmoch, U., Laville, F., Patel, P., & Frietsch, R. (2003). Linking technology areas to industrial sectors. *Final Report to the European Commission, DG Research*, *1*(0), 100.

Van Looy, B., Vereyen, C., & Schmoch, U. (2014). Patent Statistics: Concordance IPC V8–NACE Rev. 2. *Eurostat, Euopean Commission*.

Van Looy, B., Vereyen, C., & Schmoch, U. (2015). Patent Statistics: Concordance IPC V8-NACE Rev. 2 (version 2.0). *Final report, EUROSTAT, Luxembourg*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Verspagen, B., Van Moergastel, T., & Slabbers, M. (1994). MERIT concordance table: IPC-ISIC (rev. 2).

Wang, P., Fu, Y., Zhang, J., Wang, P., Zheng, Y., & Aggarwal, C. (2018). You are how you drive: Peer and temporal-aware representation learning for driving behavior analysis. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2457-2466).

Wang, P., Li, X., Zheng, Y., Aggarwal, C., & Fu, Y. (2019). Spatiotemporal representation learning for driving behavior analysis: A joint perspective of peer and temporal dependencies. *IEEE Transactions on Knowledge and Data Engineering*, *33*(2), 728-741.