WPCM: A Multi-Label Patent Classification Method Based on Weakly Supervised Learning

Dechao Wang[a], Yongjie Li[b], Jian Zhu[c] and Xiaoli Tang[d]
Institute of Medical Information & Library, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China

Keywords: Automatic Patent Classification, Weakly Supervised Learning, Multi-Label Classification, Text Feature Extraction.

Abstract: Current research on automatic patent classification predominantly focuses on reclassification within existing patent classification systems. This study aims to enhance the classification performance of automatic patent classification tasks in scenarios lacking annotated data, broaden the application scope of patent classification, and establish a foundation for mapping patents to real-world scenarios or subject-specific classification systems. To achieve this, we propose a weakly supervised multi-label patent classification method. This approach captures semantic similarity features both within patent documents and between patents and hierarchical classification labels through a two-stage process involving contrastive learning and comprehensive classification, enabling the automatic classification of unlabeled patents. Experimental results on a medical patent dataset demonstrate the efficacy of the proposed method. The model achieves Precision scores of 0.8237, 0.5743, and 0.4467 at the subclass, main group, and subgroup levels, respectively. Comparative and ablation experiments further validate the effectiveness of each component module within the method.

1 INTRODUCTION

With the rapid advancement of science and technology, the pace of technological iteration continues to accelerate. Patents, serving as crucial knowledge carriers that shape a nation's scientific and technological prowess and enhance industrial competitive advantage (Sun et al., 2024), have become indispensable instruments in technological competition. The generation of massive patent documents presents significant challenges for patent management, retrieval, analysis, and examination, concurrently imposing higher demands on the algorithms and systems underpinning automatic classification tasks (Kim et al., 2007). The efficient and precise realization of patent classification is paramount for the refined management of intellectual property rights, the improvement of patent search efficiency, and the reduction of manual examination costs. Furthermore, patent classification constitutes the foundational basis for patent information mining activities, including technology foresight analysis, patent value assessment, and the identification of innovation frontiers.

To address the challenge of automatic patent classification in scenarios lacking annotated data, this study proposes a two-stage weakly supervised learning method for multi-label classification of patents. This method, abbreviated as WPCM (Weakly-supervised Patent Classification Method), captures semantic similarity features both within patent documents and between patents and hierarchical classification labels. WPCM is designed to enable the automatic classification of unlabeled patents. Experimental validation demonstrates that the proposed WPCM method performs effectively on a medical patent dataset.

[a] https://orcid.org/0000-0002-1838-8168
[b] https://orcid.org/0009-0004-6306-8288
[c] https://orcid.org/0009-0008-8222-0280
[d] https://orcid.org/0000-0001-6946-3482

Wang, D., Li, Y., Zhu, J. and Tang, X.
WPCM: A Multi-Label Patent Classification Method Based on Weakly Supervised Learning.
DOI: 10.5220/0013708500004000

In *Proceedings of the 17th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2025) - Volume 1: KDIR*, pages 286-293

2 RELATED WORK

Patent offices worldwide must assign classification codes to each patent application to organize patents with similar characteristics within the same subdirectory. However, patent offices in different jurisdictions typically employ distinct classification systems, such as the United States Patent Classification (USPC) system, the European Classification (ECLA) system, and Japan's FI/F-TERM classification system. Currently, the predominant international patent classification systems are the International Patent Classification (IPC) and the Cooperative Patent Classification (CPC). These systems provide a scientific framework for organizing and managing patent documents. Nevertheless, with the continuous surge in patent volume and the imperative for patent data internationalization, classifying all patent documents using the IPC structure has become increasingly prevalent. Consequently, numerous researchers have developed automatic classification methods to assist in assigning IPC codes to patents (Chen et al., 2012).

Automatic patent classification refers to the process of organizing and categorizing vast quantities of patent documents without manual intervention. Contemporary research in this domain primarily encompasses three categories:

(1) Metadata-based classification methods (STUTZKI et al., 2016; Fall et al., 2003; Lim et al., 2016; Jia et al., 2017). These approaches primarily leverage structured information within patent documents, such as application numbers, applicants, classification codes, keywords, and word frequency features. However, the limited informational scope of metadata often hinders the capture of intricate details pertaining to the patent's technical content, leading to inherent accuracy limitations.

(2) Content mining-based classification methods (Jia et al., 2017; Hu et al., 2018; Bai et al., 2020; Liu et al., 2024; Li et al., 2018; Lyv et al., 2020). These methods have effectively addressed the constraints of metadata-based approaches by extracting key information from the detailed technical descriptions within patents. However, the inherent professionalism and complexity of patent texts escalate the difficulty of feature extraction and model training.

(3) Citation-enhanced classification methods (Li et al., 2009; Zhu et al., 2015; Lu et al., 2020; Peng et al., 2008). This category exploits citation relationships between patents to augment the perceived technical correlations among them, thereby aiming to improve classification accuracy. A significant research challenge, however, lies in integrating citation information into classification models due to issues concerning the availability and consistency of citation data.

Flat patent classification involves allocating patent texts to categories within the classification system at the same level using uniform criteria(GÓMEZ et al., 2019). While simpler, this approach typically yields relatively low accuracy when dealing with a vast number of patent categories and struggles to model the complex inter-category relationships. In contrast, hierarchical patent classification, building upon flat classification, incorporates the inherent hierarchical structure information of the classification system. By explicitly considering the hierarchical relationships between categories, hierarchical classification models generally achieve superior performance compared to their flat counterparts (Chen et al., 2012). Although hierarchical methods can better accommodate the hierarchical nature of patent categories (Miao et al., 2016; GÓMEZ et al., 2019), they suffer from higher computational complexity, and model performance may degrade as the number of hierarchical levels increases (Chen et al., 2012; Mun et al., 2019).

Crucially, the majority of existing patent classification methods rely heavily on large volumes of labeled data. Research on methods specifically designed for scenarios with scarce training samples or lacking annotated data remains relatively limited. Labeled data refers to text with assigned category annotations, whereas unlabeled data lacks such annotations. In practical applications, the cost of acquiring labeled data is substantial, particularly in emerging technological domains where such data may be severely scarce. Furthermore, the need for personalized classification systems in enterprise or technology management contexts−where patents may require classification according to frameworks like TRIZ theory or bespoke systems tailored to specific needs (Cong et al., 2008; Zhang et al., 2014; Hu et al., 2015; Liu et al., 2016) − presents significant challenges for conventional methods reliant on standard labeled datasets. These scenarios lacking annotated data impose new demands on classification methodologies.

Weakly supervised learning offers a promising approach, utilizing limited labeled data alongside abundant unlabeled data for model training (Liu et al., 2015). It effectively addresses challenges such as labeled data scarcity and large-scale data processing (Xiong et al., 2022). Patent classification based on weakly supervised learning does not depend on manually annotated training samples to build

classifiers. Instead, it leverages category descriptions (e.g., category names or indicative keywords) to perform classification. By fully exploiting the information within unlabeled data, weakly supervised methods can provide richer feature representations for patent classification, thereby enhancing performance. This paradigm offers distinct advantages in assisting manual classification, reducing data annotation costs, improving the accuracy and scalability of existing models, and efficiently handling large-scale datasets.

3 DATA

3.1 Data Collection

This study collected patent data pertaining to subclass A61 (Medical or Veterinary Science; Hygiene) within the IPC classification system. The classification codes and names of all hierarchical categories under the Hygiene section constituted the classification label set, comprising a total of 2,410 category labels. Concurrently, metadata for all granted U.S. invention patents classified under A61 from 2019 to 2023 were retrieved from the incoPat database. This yielded a medical patent dataset of 184,843 patent families.

3.2 Pre-Trained Language Model

PubMedBERT (Gu et al., 2022) undergoes pre-training directly on the PubMed biomedical corpus. It employs the Masked Language Modeling (MLM) mechanism to predict masked words using surrounding context and the Next Sentence Prediction (NSP) mechanism to discern sentence relationships, thereby capturing comprehensive semantic vector representations and robust global text features relevant to biomedicine (Dong et al., 2021). PubMedBERT incorporates domain-specific optimizations, including filtering non-medical corpora and integrating additional medical terminology, which enhance its performance for biomedical applications (Gu et al., 2022). Given that the patent data collected in this study is confined to IPC subclass A61 (Medical/Veterinary Science; Hygiene), PubMedBERT was selected as the pre-trained language model.

3.3 Label Vectorization

The weakly supervised patent classification method leverages, beyond patent metadata, only the textual descriptions of classification categories and their inherent hierarchical information. Categories within the classification label set are structured hierarchically (Shen et al., 2023), where lower-level categories are semantically constrained by their ancestors (Rojas et al., 2020), and higher-level categories encapsulate the scope of their descendants.

To generate vector representations for all category names in the label set, PubMedBERT is utilized. Subsequently, a hierarchical weighting scheme, informed by the category structure, is applied to the vectors at different levels. This ensures that the final category vectors reflect the specific content of descendant categories while remaining semantically constrained by their ancestors.

The hierarchical processing of classification labels is illustrated in Figure 1.

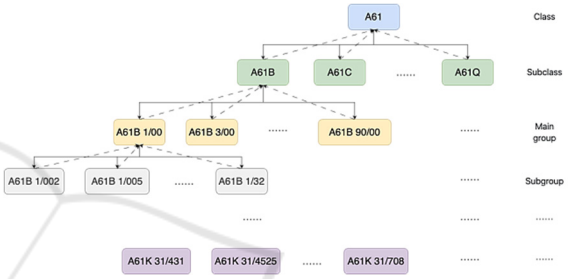

Figure 1: Hierarchical process of classification labels.

Categories situated between the root and leaf levels are influenced by both their parent and child categories. The final vector representation El for a category label l is derived from the weighted sum and average of three components:

(1) The vector el generated directly by PubMedBERT based on the category name l.

(2) The average vector of all immediate child categories of l (if any).

(3) The average vector of all ancestor categories of l tracing upward to the root category (if any).

This process is formalized in Formula (1):

$$E_{A61B\ 1/00} = \frac{1}{3}\left(e_{A61B\ 1/00} + \frac{\left(e_{A61B\ 1/002} + e_{A61B\ 1/005} + \cdots + e_{A61B\ 1/32}\right)}{n_1} + \frac{\left(e_{A61B} + e_{A61}\right)}{n_2}\right) \quad (1)$$

Where:

El is the final vector representation of category label l.

ek represents the vector of category k generated by the pre-trained language model.

n1 denotes the number of direct child categories under l.

n2 denotes the number of ancestor categories from l up to the root category.

4 METHODS

4.1 Research Framework

This study proposes a two-stage weakly supervised learning method for multi-label patent classification (WPCM). In the first stage, the pre-trained language model (PubMedBERT) is fine-tuned by capturing semantic features intrinsic to patents. In the second stage, classification is performed based on the semantic similarity between patent representations and hierarchical classification label vectors. The overall workflow is illustrated in Figure 2.

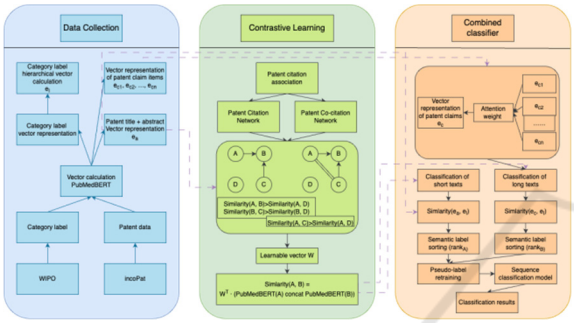

Figure 2: The multi-label patent classification method based on weakly supervised learning.

4.2 Contrastive Learning Stage

Common references among patents reflect the similarity of the underlying technologies they represent (Lai et al., 2005). Co-citation analysis facilitates the identification of patent clusters and their interrelationships (Peng et al., 2008). Leveraging this, citation networks and co-citation networks are constructed from the medical patent dataset. Given that connected nodes within these networks exhibit higher similarity than unconnected pairs, a contrastive learning strategy is designed to fine-tune PubMedBERT. This strategy aims to maximize the similarity between patent representations with citation/co-citation links ((pi, pj)) while minimizing the similarity between unlinked pairs ((pi, pk)). The specific formulation is detailed in Formulas 2 to 4.

$$e_{pos1} = \text{PubMedBERT}(pt_i) \text{ concat } \text{PubMedBERT}(pt_j) \quad i, j \in (1, n) \tag{2}$$

$$e_{neg1} = \text{PubMedBERT}(pt_i) \text{ concat } \text{PubMedBERT}(pt_k) \quad i, k \in (1, n) \tag{3}$$

$$\text{sim}(p_i, p_j) - \text{sim}(p_i, p_k) = w^T e_{pos1} - w^T e_{neg1} \tag{4}$$

Where, pti (i∈(1, n)) is the title and abstract text of the patent pi, n is the total number of patents, and w is the learnable vector. The patent pair (pi, pj) has a citation or co-citation relationship, but the patent pair (pi, pk) does not hold such correlations. Based on the research of Zhang et al. (2023), the loss function is designed as formula 5.

$$L = -\log \frac{\exp(\text{sim}(p_i, p_j))}{\exp(\text{sim}(p_i, p_j)) + \exp(\text{sim}(p_i, p_k))} \tag{5}$$

4.3 Comprehensive Classification Stage

4.3.1 Short-Text Classification (Title & Abstract)

Semantic similarity between the patent's title/abstract summary (pti, i ∈ [1, n]) and the description text (ltj, j ∈ [1, s], s = total labels) of each classification label lj is computed. Utilizing the fine-tuned vector w obtained from Section 4.2, the similarity sim(pti, ltj) is calculated according to Formula (6). This generates a semantic label ranking rankA(pi) for each patent pi.

$$\text{sim1}(p_i, l_j) = w^T(\text{PubMedBERT}(pt_i) \text{ concat } \text{PubMedBERT}(lt_j)) \tag{6}$$

4.3.2 Long-Text Classification (Claims)

Recognizing that patent titles and abstracts convey limited information, and acknowledging that claims provide richer technical content (Lee et al., 2020) with varying importance across individual entries, this study employs a Dot-Product Attention Mechanism (Bai et al., 2020) to generate weighted claim embeddings.

Specifically, the pre-trained language model first generates embeddings [ei1, ei2, ..., eim] for each term within the claims patent pi. These embeddings are then multiplied by a learnable attention vector v to compute attention weights αi for each claim term position. Finally, a weighted sum of the term embeddings yields the comprehensive claim embedding Ei for patent pi:

The similarity sim(Ei, ltj) between the claim embedding Ei and the label description ltj is calculated using Formula (7), producing the claim-based similarity ranking rankC(pi).

$$\text{sim2}(p_i, l_j) = w^T(E_i \text{ concat } \text{PubMedBERT}(lt_j)) \tag{7}$$

4.3.3 Pseudo-Label Retraining

For each patent p and potential label l, a comprehensive ranking score rank(l|p) is computed as the average reciprocal rank (MRR) of rankA(l|p) and rankC(l|p).

Pseudo-labels, automatically generated labels derived from model outputs, are assigned to the unlabeled patent data to create an enhanced training set (Mekala et al., 2020; Zhang et al., 2022). Based on rank(l|p), the top-5 classification labels with the highest scores for each patent are selected as pseudo-labels.

These pseudo-labels, combined with the patent's feature representations (title/abstract embedding and claim embedding Ei) and key metadata (e.g., number of claims, number of independent claims), are used to retrain a sequence classification model AutoModelForSequenceClassification (Uribe et al., 2022). This supervised retraining leverages the pseudo-labeled data to further refine the classification capability.

The final prediction results are obtained by correcting the initial comprehensive ranking rank(l|p) using the outputs of the retrained model. The top-n classification labels (where n equals the actual number of categories assigned to the patent) are selected as the predicted categories.

5 RESULTS

5.1 Model Evaluation

Given that the WPCM method operates without labeled data, no distinct training/test set split was performed; the entire medical patent dataset served as the evaluation corpus. For assessing the multi-label classification performance, we employed standard metrics: Normalized Discounted Cumulative Gain NDCG@k (Wang et al., 2013), Recall, Precision, and F1-score. Table 1 presents the evaluation results of WPCM across the three hierarchical classification levels (subclass, main group, subgroup). Notably, despite utilizing no labeled data, WPCM demonstrates robust classification capability.

Table 1: Test results of the algorithm.

	NDCG	Recall	Precision	F1
subclasses	0.8770	0.7912	0.8237	0.7944
main groups	0.6228	0.5247	0.5743	0.5322
subgroups	0.5050	0.3990	0.4467	0.4042

5.2 Comparative Experiments

To validate the efficacy of the contrastive learning-based feature recognition module (Stage 1 fine-tuning), we integrated this feature extraction approach into conventional supervised machine learning models. Specifically, Support Vector Machine (SVM), XGBoost, and Softmax Regression (single-layer neural network) models were trained on embedding vectors generated by PubMedBERT, using the true hierarchical classification labels for the main group level. Comparative results are shown in Table 2. The findings indicate that incorporating contrastive learning-derived features significantly enhances the classification performance of all supervised baselines.

Table 2: Comparison of experimental results.

	NDCG	Recall	Precision	F1
SVM	0.5953	0.2193	0.2769	0.2273
SVM+ Contrastive learning	0.6009	0.2236	0.3128	0.2496
XGBoost	0.3850	0.0311	0.2170	0.0478
XGBoost+ Contrastive learning	0.4454	0.0520	0.2810	0.0784
Single-layer neural Network (Softmax)	0.5960	0.1193	0.2805	0.1526
Single-layer neural network + contrastive learning	0.5643	0.4430	0.5111	0.4534
WPCM(Unlabeled Data)	0.6228	0.5247	0.5743	0.5322

5.3 Ablation Studies

Three ablation studies were conducted on the main group dataset to evaluate the contribution of each module within the Stage 2 comprehensive classification framework:

Ablation Model 1 (Contrastive Learning + Long-Text Classification + Retraining): Utilized only claim embeddings (rankC) for pseudo-label generation and retraining. The short-text classification module (title/abstract) was removed.

Ablation Model 2 (Contrastive Learning + Short-Text Classification + Retraining): Utilized only title/abstract embeddings (rankA) for pseudo-label generation and retraining. The long-text classification module (claims with attention) was removed.

Ablation Model 3 (Contrastive Learning + Short-Text + Long-Text Classification): Utilized both rankA and rankC to compute the comprehensive ranking rank(l|p) without subsequent pseudo-label retraining.

The performance of the full WPCM model versus these ablation variants is presented in Table 3.

Table 3: Results of ablation experiment.

	NDCG	Recall	Precision	F1
WPCM	0.6228	0.5247	0.5743	0.5322
Ablation Experiment 1	0.6148	0.5091	0.5744	0.5239
Ablation Experiment 2	0.6138	0.5086	0.5702	0.5225
Ablation Experiment 3	0.5960	0.4783	0.5602	0.4938

*WPCM: Contrastive Learning + Short Text Classification + Long Text Classification + Retraining. Ablation Experiment 1: Contrastive learning + long text classification + retraining. Ablation Experiment 2: Contrastive learning + Short text classification + retraining. Ablation Experiment 3: Contrastive learning + Short text classification + long text classification.

WPCM outperformed all ablation models across key metrics:

Compared to Ablation Model 1: NDCG increased by 0.80 percentage points (pp), Recall by 1.56 pp, F1-score by 0.83 pp.

Compared to Ablation Model 2: NDCG increased by 0.90 pp, Recall by 1.61 pp, F1-score by 0.97 pp.

Compared to Ablation Model 3: NDCG increased by 2.68 pp, Recall by 4.64 pp, F1-score by 3.84 pp. Precision increased by 1.41 pp.

These results confirm that the integrated comprehensive classification method—leveraging both patent text modalities and pseudo-label refinement—effectively enhances classification accuracy by capturing semantic similarities between patent features and hierarchical label vectors.

Furthermore, the superior performance of Ablation Model 1 (retaining claims) over Ablation Model 2 (retaining title/abstract) provides empirical support for the established view that patent claims offer richer and more representative technical content than titles and abstracts alone (Lee et al., 2020).

6 DISCUSSION

This study proposes WPCM, a weakly supervised learning method for multi-label patent classification, designed to address the challenges of automatic patent classification in scenarios with scarce or entirely missing labeled data. Empirical evaluation on the medical patent dataset demonstrates that WPCM achieves robust performance without utilizing labeled training data. This approach significantly reduces data annotation costs and offers valuable assistance to patent examiners. Furthermore, WPCM's text feature extraction methodology—particularly its handling of semantic similarity features—can be adapted to enhance other patent classification models, potentially improving their performance and generalizability.

Unlike conventional supervised patent classification models requiring large-scale labeled datasets, WPCM's core innovation lies in its weakly supervised framework, enabling effective classification in unlabeled scenarios. Additionally, the integration of contrastive learning leveraging citation relationships refines feature recognition, allowing the model to capture nuanced semantic associations between patent texts and hierarchical classification labels more accurately. This not only boosts classification performance but also provides an effective feature recognition strategy applicable to existing supervised models. WPCM's multi-stage architecture offers a novel paradigm for text feature mining in complex classification tasks.

The core challenge in automatic patent classification persists: effectively processing massive patent volumes while accommodating the complexity of patent texts and the hierarchical nature of classification systems. Future research must focus on enhancing the accuracy of semantic feature extraction and optimizing classifier fusion strategies to improve model applicability and generalization.

Despite its strengths, WPCM has limitations:

Domain Specificity: Validation was conducted solely on the medical patent dataset. Future work must assess WPCM's generalization capability across diverse technical domains and classification systems.

Data Dependency: WPCM's training requires sufficient textual features (titles, abstracts, claims). Performance under conditions of severely scarce or low-quality data warrants further investigation.

Hierarchical Modeling: While WPCM utilizes hierarchical label information during vectorization (Section 3.3), it does not explicitly incorporate hierarchical classification during model prediction. This may lead to suboptimal utilization of structural information, potentially impacting performance.

Future research will:

Evaluate WPCM's performance across diverse technical domains and classification systems (e.g., CPC, USPC) to assess its generalization scope.

Explore integrating domain-specific knowledge bases or specialized pre-trained language models to enhance semantic understanding via external knowledge, thereby boosting domain-adaptive classification capability.

Investigate explicit hierarchical classification mechanisms within the WPCM framework to better leverage the structure inherent in patent classification systems.

REFERENCES

Bai, J.H., Shim, I.S., & Park, S.H. (2020). MEXN: Multi-stage extraction network for patent document classification. Applied Sciences, 10(7), 2472.

Chen, Y.L., & Yuan, C. (2012). A three-phase method for patent classification. Information Processing & Management, 48, 1017–1030.

Cong, H., & Loh, H.T. (2008). Grouping of TRIZ inventive principles to facilitate automatic patent classification. Expert Systems with Applications, 34, 788–795.

Dong, M., Su, Z.Q., Zhou, X.B., et al. (2021). Improving PubMedBERT for CID-entity-relation classification using Text-CNN [In Chinese]. Data Analysis and Knowledge Discovery, 5(11), 145–152.

Fall, C.J., Törcsvári, A., Benzineb, K., & Karetka, G. (2003). Automated categorization in the international patent classification. SIGIR Forum, 37(1), 10–25.

Gómez, J.C. (2019). Analysis of the effect of data properties in automated patent classification. Scientometrics, 121, 1239–1268. https://doi.org/10.1007/s11192-019-03246-1

Gu, Y., Tinn, R., Cheng, H., et al. (2022). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare, 3(1), 1–23.

Hu, J., Li, S., Hu, J.Y., & Yang, G. (2018). A hierarchical feature extraction model for multi-label mechanical patent classification. Sustainability, 10(3), 1–22.

Hu, Z.Y., Fang, S., Wen, Y., et al. (2015). Information extraction based on functional basis and experimental study on automatic classification [In Chinese]. Data Analysis and Knowledge Discovery, (1), 66–74.

Jia, S.S., Liu, C., Sun, L.Y., et al. (2017). Patent classification based on multi-feature and multi-classifier integration [In Chinese]. Data Analysis and Knowledge Discovery, 1(8), 76–84.

Kim, J.H., & Choi, K.S. (2007). Patent document categorization based on semantic structural information. Information Processing & Management, 43, 1200–1215.

Lai, K.K., & Wu, S. (2005). Using the patent co-citation approach to establish a new patent classification system. Information Processing & Management, 41, 313–330.

Lee, J.S., & Hsiang, J. (2020). Patent classification by fine-tuning BERT language model. World Patent Information, 61, 101965.

Li, S., Hu, J.Y., Cui, Y.F., & Hu, J. (2018). DeepPatent: Patent classification with convolutional neural networks and word embedding. Scientometrics, 117, 721–744.

Li, X., Chen, H.J., Zhang, Z.W., et al. (2009). Managing knowledge in light of its evolution process: An empirical study on citation network-based patent classification. Journal of Management Information Systems, 26(3), 129–154.

Lim, S., & Kwon, Y.J. (2016). IPC multi-label classification based on the field functionality of patent documents. In J. Li et al. (Eds.), Advanced Data Mining and Applications (pp. 1–12). Springer.

Liu, K., Peng, S.W., Wu, J.Q., et al. (2015). MeSHLabeler: Improving the accuracy of large-scale MeSH indexing by integrating diverse evidence. Bioinformatics, 31(19), i339–i347.

Liu, L.F., Li, Y., Hou, C.Y., et al. (2016). Information extraction based on functional basis and experimental study on automatic classification [In Chinese]. Advanced Engineering Sciences, 48(5), 105–113. https://doi.org/10.15961/j.jsuese.2016.05.016

Liu, Y., Xu, F., Zhao, Y., et al. (2024). Hierarchical multi-instance multi-label learning for Chinese patent text classification. Connection Science, 36.

Lu, Y.H., Xiong, X.W., Zhang, W.J., et al. (2020). Research on classification and similarity of patent citation based on deep learning. Scientometrics, 123, 813–839.

Lyu, L.C., Han, T., Zhou, J., et al. (2020). Research on the method of Chinese patent automatic classification based on deep learning [In Chinese]. Library and Information Service, 64(10), 75–85. https://doi.org/10.13266/j.issn.0252-3116.2020.10.009

Mekala, D., Zhang, X., & Shang, J. (2020). META: Metadata-empowered weak supervision for text classification. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (pp. 8351–8361).

Miao, J.M., Jia, G.W., & Zhang, Y.L. (2016). Research on the method of Chinese patent automatic classification based on deep learning [In Chinese]. Information Studies: Theory & Application, 39(8), 103–105, 91. https://doi.org/10.16353/j.cnki.1000-7490.2016.08.020

Mun, C., Yoon, S., & Park, H. (2019). Structural decomposition of technological domain using patent co-classification and classification hierarchy. Scientometrics, 121, 633–652.

Peng, A.D. (2008). Research on patent classification method and related problems based on co-citation [In Chinese]. Information Science, (11), 1676–1679, 1684.

Rojas, K.R., Bustamante, G., Sobrevi lla, M.A.C., & Oncevay, A. (2020). Efficient strategies for hierarchical text classification: External knowledge and auxiliary tasks. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 2252–2257).

Shen, J.H., Chen, H.Y., Zhang, G.P., et al. (2023). Research on patent text classification model based on hierarchical classifier [In Chinese]. Journal of Intelligence, 42(8), 157–163, 68.

Stutzki, J., & Schubert, M. (2016). Geodata supported classification of patent applications. Proceedings of the Third International ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data.

Sun, C.H., Cai, H.Y., Shi, J., et al. (2024). Research on identifying key core technologies from the perspective of patent application [In Chinese]. Information Studies: Theory & Application. Advance online publication. http://kns.cnki.net/kcms/detail/11.1762.G3.20240824.1528.002.html

Uribe, D., Cuan, E., & Urquizo, E. (2022). Fine-tuning of BERT models for sequence classification. 2022 International Conference on Mechatronics, Electronics and Automotive Engineering (pp. 140–144).

Wang, Y., Wang, L., Li, Y., et al. (2013). A theoretical analysis of NDCG type ranking measures. Proceedings of the 26th Annual Conference on Learning Theory (pp. 1–10).

Xiong, Y., Chang, W.C., Hsieh, C.J., et al. (2021). Extreme zero-shot learning for extreme text classification. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (pp. 5455–5468).

Zhang, L.F., Shah, S.K., & Kakadiaris, I.A. (2017). Hierarchical multi-label classification using fully associative ensemble learning. Pattern Recognition, 70, 89–103.

Zhang, X.Y. (2014). Interactive patent classification based on multi-classifier fusion and active learning. Neurocomputing, 127, 200–205.

Zhang, Y., Jin, B.W., Chen, X.S., et al. (2023). Weakly supervised multi-label classification of full-text scientific papers. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

Zhang, Y., Shen, Z.H., Wu, C.H., et al. (2022). Metadata-induced contrastive learning for zero-shot multi-label text classification. Proceedings of the ACM Web Conference 2022 (pp. 1–10).

Zhu, F.J., Wang, X.W., Zhu, D.M., & Liu, Y.F. (2015). A supervised requirement-oriented patent classification scheme based on the combination of metadata and citation information. International Journal of Computational Intelligence Systems, 8(3), 502–516.