arXiv:2508.15800v1 [cs.CL] 13 Aug 2025

A BERT-based Hierarchical Classification Model with Applications in Chinese Commodity Classification

Kun Liu, Tuozhen Liu, Feifei Wang, and Rui Pan

[a]*School of Statistics and Mathematics, Central University of Finance and Economics, 2023210967@email.cufe.edu.cn, Beijing, 100081, Beijing, China*
[b]*Reigning Capital Co., Ltd., liutuozhen@reigning-capital.com, Beijing, 100080, Beijing, China*
[c]*Center for Applied Statistics and School of Statistics, Renmin University of China, feifei.wang@ruc.edu.cn, Beijing, 100872, Beijing, China*
[d]*School of Statistics and Mathematics, Central University of Finance and Economics, ruipan@cufe.edu.cn, Beijing, 100081, Beijing, China*

Abstract

Existing e-commerce platforms heavily rely on manual annotation for product categorization, which is inefficient and inconsistent. These platforms often employ a hierarchical structure for categorizing products; however, few studies have leveraged this hierarchical information for classification. Furthermore, studies that consider hierarchical information fail to account for similarities and differences across various hierarchical categories. Herein, we introduce a large-scale hierarchical dataset collected from the JD e-commerce platform (*www.JD.com*), comprising 1,011,450 products with titles and a three-level category structure. By making this dataset openly accessible, we provide a valuable resource for researchers and practitioners to advance research and applications associated with product categorization. Moreover, we propose a novel hierarchical text classification approach based on the widely used Bidirectional Encoder Representations from Transformers (BERT), called Hierarchical Fine-tuning BERT (HFT-BERT). HFT-BERT leverages the remarkable text feature extraction capabilities of BERT, achieving prediction performance comparable to those of existing methods on short texts. Notably, our HFT-BERT model demonstrates exceptional performance in categorizing longer short texts, such as books.

Keywords: BERT, hierarchical text classification, fine-tuning, product categorization

1. Introduction

Product categorization plays a crucial role in ensuring effective functioning of e-commerce platforms [55, 10]. It involves categorizing products based on their nature, composition, and intended use, allowing for efficient search and filtering mechanisms. This categorization process is the backbone of product organization [13]. Through an effective classification mechanism, consumers can easily find what they need, while sellers can effectively showcase their products. However, existing e-commerce platforms heavily rely on manual product annotation [17]. According to the rules of various e-commerce platforms for adding products, sellers must manually select or fill in product categories [20]. However, the manual approach has low efficiency and consistency because various sellers might categorize the same product differently [55].

1.1. Problem Statement

In e-commerce platforms, sellers are required to choose the most appropriate product categories from the numerous categories available [5]. Therefore, scholars have adopted text classification methods to classify products based on their titles. Specifically, Lee et al. [29] adapt the doc2vec algorithm that implements the document embedding technique to develop an automatic product classifier. Chen et al. [5] propose a neural product categorization model to identify fine-grained categories from the product content. A character-level convolutional embedding layer is used to learn compositional word representations, and a spiral residual layer is used to extract word context annotations that capture complex long-range dependencies and structural information. [32] propose a text-image adaptive convolutional neural network to effectively utilize website information and facilitate the customs classification process. Later, Zhang et al. [55] propose a commodity text-classification-based e-commerce category and attribute mining method, which takes into account the attributes of the corresponding categories between different platforms. Recently, Fang et al. [13] use BERT to obtain text feature representations and a Variational Auto-Encoder (VAE) to address title discrimination and unbalanced sample sizes. This integrated approach is referred to as BERT-VAE.

Notably, products on e-commerce platforms often have a hierarchical category structure [27]. To the best of our knowledge, most existing literature utilizes the technique of flat classification, ignoring the hierarchical structure

[54, 33]. Recently, methods have been proposed for hierarchical classification, enhancing the predictive ability of text classification models [14, 16, 36, 46]. However, many existing studies focus on utilizing direct parent–child relationships in the hierarchy, ignoring differences and similarities in the same hierarchy level [46]. Further, most applications of the existing hierarchical text classification (HTC) methods deal with English texts, such as the RCV1 dataset [31] and the Amazon670K dataset [30]. To the best of our knowledge, few hierarchical classification methods have been developed for Chinese texts. Therefore, developing a method that can make the full use of hierarchical information and is suitable for various hierarchical structures as well as Chinese texts is crucial.

1.2. Research Objective

To make full use of hierarchical information, the Hierarchical Fine-Tuning based Convolutional Neural Network (HFT-CNN) proposed by Shimura et al. [42] effectively utilizes the hierarchical information of categories and achieves better prediction results. Data in the upper levels are used for categorization in the lower levels by applying a CNN with a fine-tuning technique. Inspired by them [42], we use the BERT model instead of a CNN for the hierarchical classification of product texts. Because of the better performance of BERT in text feature extraction and Chinese text learning, we aim to achieve more accurate classification results through this replacement. Using the pretrained model of BERT, we can learn contextual information in product texts and apply it for hierarchical classification. Compared with CNN, BERT can better capture semantics and contextual information in texts, improving classification accuracy. We fine-tune the BERT model to adapt it to the hierarchical classification of product texts using its pretrained weights as the starting point. With this improvement, we obtain better prediction results and enhance the performance of hierarchical classification of product texts.

1.3. Contribution

Based on the above discussions, we present the contributions of our paper from three perspectives. First, we collect a large-scale Chinese dataset of products from the JD e-commerce platform (*www.JD.com*), which comprises 1,011,450 products. This dataset includes product titles and category names at three hierarchical levels in Chinese. There is no space between words and no punctuation mark after word endings in the Chinese text. Moreover, the number of high-frequency words in Chinese is substantially larger than that in English. Therefore, the analysis of the Chinese text is more difficult [34]. Existing methods mostly applied to English text, while our work is based on Chinese text classification. In addition, by making this dataset open-source (*https://gitee.com/KunLiu_kk/jd-dataset*), we provide a valuable resource for researchers and practitioners to advance research and applications associated with product categorization. Second, we propose a novel approach for HTC that incorporates the extensively used BERT model and utilizes hierarchical information in the dataset, referred to as Hierarchical Fine-Tuning BERT (HFT-BERT). This approach leverages the superior text feature extraction capabilities of BERT, achieving prediction performance comparable to those of existing methods on our dataset. Third, the HFT-BERT model shows short-text classification performance comparable to those of other existing models and achieves excellent prediction performance on longer short texts, e.g., book introductions, compared to those of other existing models. Thus, our model can provide support for more efficient operations on Chinese e-commerce platforms.

2. Related Work

HTC plays an important role in text classification, where documents are organized in a hierarchical structure. In the early years, the most commonly employed technique to solve the HTC problem is the local classifier approach, which mainly constructs a hierarchy of traditional classifiers [49]. According to Silla et al. [43], the local classifier approach can be further categorized into three standard methods: training a local classifier per node (LCN), training a local classifier per parent node (LCPN), and training a local classifier per level (LCL). In contrast to the local classifier approach, the global classifier approach first treats HTC problem as a flat multiclassification problem and uses a single classifier [57]. Hierarchical information is introduced into the global classifier approach using special parameter initialization, regularization term, or model structure. Methods developed for multi-label text classification problems can be subsequently used [52, 2, 48].

2.1. Traditional Classifier

In the LCN approach, a classifier is trained for each category in the hierarchy to determine whether an item should be classified into this category. The classifiers for each category can be the same or different. For instance, D'Alessio et al. [7] utilized a one-of-M classifier at the top level, along with binary classifiers for each category in sub trees. To make better use of the parent–child relationship in the hierarchy, Sun et al. [47] further built a support vector machine (SVM) classifier for each category and a binary classifier for each parent category. In contrast to considering the tree structure, Nguyen et al. [38] proposed a technique based on directed acyclic graphs, where a child category may have multiple parent categories. Here, the SVM is used as a binary classifier for each category. Thus, novel structures, new classifiers, and algorithms continuously emerge in relation to the LCN approach [12, 11, 18].

In the LCPN approach, each parent category is associated with a multilabel classifier or a series of binary classifiers to determine which child category or categories an item should be classified into [43]. Koller et al. [24] pioneered an HTC solution using the LCPN approach, where Naive Bayes was selected as the baseline classifier for each parent category. Wu et al. [51] utilized the C4.5 decision tree as the base multilabel classifier for each parent category. Further, Moskovitch et al. [37] employed centroid classifiers to determine whether an item should be assigned to corresponding categories based on a threshold. Typically, the same classification algorithm is used throughout the hierarchy in the LCPN approach. However, Secker et al. [41] used different classification algorithms at different parent nodes of the class hierarchy to improve prediction accuracy. Secker et al. [40] further extended the approach proposed in 2007 [41] by selecting different classifiers as well as attributes at each step when choosing classifiers.

In the LCL approach, a multilabel classifier is trained for each level in the hierarchy to determine which category or categories at that level an item should be classified into [43]. Clare et al. [6] were the first to employ the LCL method and used the C4.5 decision tree as their multilabel classifier. Chen et al. [3] used the back propagation (BP) learning model to construct appropriate hierarchical classification units for each level, which constitutes an LCL approach. In this approach, the results from top level trigger the BP classifiers in the next level, and so on.

Unlike the local classifier, the global classifier uses a single classifier, treating the HTC problem as a flat multiclassification problem with special hierarchical information. The global classifier approach is also referred to as the Big-Bang approach [43]. Different traditional classifiers are used as the multilabel global classifiers [15, 8, 44]. In addition to using different global classifiers, several studies focus on optimizing models and improving the effectiveness of hierarchical classification. Kiritchenko et al. [23] were the first to incorporate the boosting algorithm into HTC methods, presenting a novel global classifier approach based on AdaBoost, which achieved consistent classification results. Khan et al. [22] proposed a novel ant-colony-optimization-based single-path hierarchical classification algorithm, which was further extended in their later research [21].

2.2. Deep Learning Classifier

Deep learning algorithms have attracted considerable attention since their emergence and have been widely applied in various fields particularly in text analysis. Recently, there has been a surge in research exploring the applications of deep learning algorithms to sovle HTC problems. Lecun et al. [28] demonstrated the effectiveness of deep learning models in automatically learning the hierarchical representations of image data, leading to the widespread adoption of CNNs in the field of HTC [39, 14]. These CNN-based global approaches utilize graph convolution operations to process texts represented as a graph-of-words. In addition to the CNN model, other deep learning models such as encoder and decoder models have also been introduced into HTC methods to extract textual and label semantics [16, 4, 35, 53]. However, the training models for global classifier approaches are logically complicated and often suffer from underfitting owing to lost priori information about categories and their structural relationships [26]. In contrast, local classifier approaches can make better use of hierarchical information. Therefore, many studies combine deep learning with local classifier approaches.

Regarding local classifier approaches, Kowsari et al. [26] were the first to propose a local HTC approach called HDLTex based on deep learning models, utilizing Deep Neural Networks (DNNs), CNNs, and Recurrent Neural Networks (RNNs) as multilabel classifiers. The HDLTex model builds a multilabel classifier for elements belonging to the same parent category. The authors constructed a hierarchical two-level taxonomy dataset of 46,985 published papers obtained from the Web of Science *www.webofscience.com.* The results showed that RNNs exhibited the best classification accuracy, followed by the CNN model. Based on transfer learning methods, Banerjee et al. [1] proposed HTrans, where binary classifiers at lower levels in the hierarchy are initialized using the parameters of the parent classifier and fine-tuned during the child-category classification task. Transfer learning methods seem to offer better performance in solving HTC problems. Compared with the attention-based gated recurrent unit (GRU) model, the HTrans method offers significant improvements of 1% concerning micro-F1 and 3% concerning macro-F1 on RCV1. However, the above-mentioned methods involve a huge number of parameters and computational overhead because they build or fine-tune deep learning models for almost every node [45]. Wang et al. [50] propose a novel hierarchical classification method based on graph learning model is proposed to learn the graph embedding that well captures the node, relation, and graph structure information for hierarchical classification.

Among various approaches that utilize deep learning, the LCL approach has gained significant popularity. In the LCL approach, deep networks are constructed or fine-tuned locally for each level within the hierarchy. This strategy effectively addresses the challenges of excessive parameters and computational overhead. Shimura et al. [42] proposed the HFT-CNN, a CNN-based LCL method with a fine-tuning technique. Results based on the RCV1 dataset [31] and the Amazon670K dataset [30] prove the competitiveness of fine-tuning in solving HTC problems. Meanwhile, Sinha et al. [45] proposed an attention-based Long Short-Term Memory (LSTM) encoder model using the WOS dataset [26]. They showed that the use of hierarchical taxonomy can provide a more robust classifier than flat classifiers. Kowsari wt al. [26] reported that the RNN model has the best effect; hence, RNNs are popular models used in the LCL approaches. Based on the RNN model, Huang et al. [19] proposed Hierarchical Attention-based RNN (HARNN), an LCL approach that develops a hierarchical attention-based recurrent layer to capture associations between texts and the hierarchical structure. Similarly, Ma et al. [36] proposed a hybrid embedding-based text representation for hierarchical multi-label text classification (HE-HMTC), wherein a Bidirectional GRU (Bi-GRU) model is trained at each level of the category hierarchy or taxonomy in a top–down manner. The HE-HMTC model performs well on the WOS [26] and other datasets. However, the relationship between levels remains underutilized in HARNN and HE-HMTC. To share information across hierarchy levels more flexibly and effectively, Zhang et al. [56] designed a novel label-based attention module, which can hierarchically extract important information from the text based on labels from different hierarchy levels.

Among the LCL approaches described above, HFT-CNN [42] is the most similar to our method. HFT-CNN utilizes data in the upper levels for categorization in the lower levels by applying a CNN model with a fine-tuning technique. In other words, HFT-CNN transfers the parameters of CNN trained from the upper to lower levels according to the hierarchical structure and fine-tunes the parameters. The basic idea of fine-tuning is to adopt a pretrained model that has been trained on a large number of texts and continue to train it on a small number of texts [9]. As a result, the fine-tuning technique is suitable for HTC because the sample size is smaller for the lower level in the hierarchical structure. Furthermore, hierarchical dependencies between labels lead to similar parameters, making the fine-tuning of parameters effective [39, 25]. Therefore, instead of training a CNN model on the top level, we select the pretrained BERT model called BERT-base-Chinese and fine-tune the parameters on each level in the hierarchical structure from the top to bottom.

3. HFT-BERT framework

The framework of our model is shown in Figure 1. HFT-BERT is specifically designed for hierarchical classification, and uses the BERT-base-Chinese model as its foundation. The BERT-base-Chinese model, developed by HuggingFace team (*https://huggingface.co/bert*), is a pretrained model catering to simplified Chinese languages. This model is built upon a BERT-based architecture, which was initially proposed by Devlin et al. [9]. Within our HFT-BERT framework, the initial parameters of the BERT-base-Chinese model are imported from the pretrained model, as detailed in *https://huggingface.co/bert*. These parameters are updated during the training process at the initial level. Then, the updated parameters at the current level are transferred to the subsequent level, where they undergo further fine-tuning during training. In addition to the BERT-base-Chinese model, each level within the HFT-BERT architecture incorporates a DropOut layer and fully connected layer with a Softmax activation function. Notably, the parameters from these two layers are also updated during training at each level within the hierarchical structure. However, they are not transferred or fine-tuned across different levels, maintaining their level-specific independence. The inclusion of the DropOut layer is a crucial step after using the BERT-base-Chinese model to avoid overfitting. Subsequently, a fully connected layer featuring a Softmax activation function is introduced as the classifier. The Softmax activation function calculates the probabilities associated with products belonging to each category within the current level. Finally, the prediction of the category of a product is determined by selecting the category with the highest probability.

To facilitate the implementation of the BERT-base-Chinese model, texts

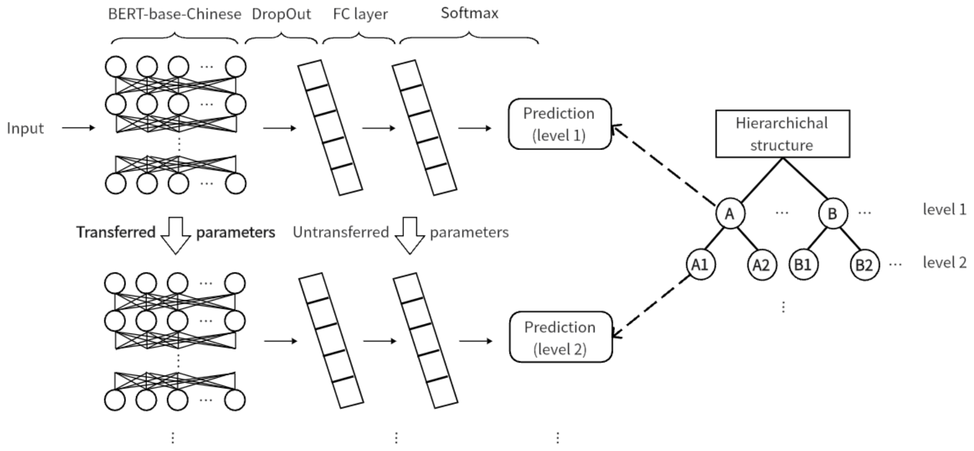

Figure 1: Framework of HFT-BERT. The initial parameters of the BERT-base-Chinese model are loaded from the pretrained model and transferred and fine-tuned between different levels. Each level also has its own DropOut layer to avoid overfitting and a fully connected layer with the Softmax activation function as a classifier.

that do not reach the maximum length are padded with the special character "<pad >" during formatting. The HFT-BERT model operates sequentially, moving from the top to the bottom, indicating that it processes levels from 2 to 3 within the hierarchical structure of our dataset. As a result, at level 2, categories are encoded and fed into the BERT-base-Chinese model for a 12-layer bidirectional Transformer operation, together with the word embedding vectors of the formatted texts. The parameters of the BERT-base-Chinese model are initially sourced from the pretrained BERT-base-Chinese model provided by HuggingFace team (*https://huggingface.co/bert*). Subsequently, the outputs generated by the BERT-base-Chinese model are directed into the DropOut layer and the fully connected layer with a Softmax activation function. We employ the cross-entropy loss function to perform gradient back propagation. Parameters are updated using the cosine-annealing learning rate schedule. During the training and evaluation phases, a batch size of 128 is employed. Upon completing the training and evaluation procedures at level 2, the model is then assessed using the test set. The prediction accuracy for level 2 is computed on this test set.

As HFT-BERT moves to level 3, a noteworthy aspect is the sharing of parameters between levels 2 and 3. The parameters that were fine-tuned during the operation at level 2 are loaded as the initial parameters for level 3, and they continue to undergo fine-tuning. At this stage, categories specific to level 3 are encoded and introduced into the BERT-base-Chinese model for a 12-layer bidirectional Transformer operation, along with the word embedding vectors of the formatted texts. Notably, the parameters within the DropOut layer and the fully connected layer are initialized separately and are not shared across levels. Following the same sequence of steps employed at level 2, the test set is utilized, and the prediction accuracy for level 3 is determined.

4. Experiments and Results

4.1. Data Description

We evaluate the proposed model using a real-world dataset characterized by a three-layer hierarchical structure. The dataset is obtained from JD.com, a prominent supply-chain-focused technology and service provider and a well-known comprehensive e-commerce platform in China. Our dataset comprises four categories in level 1:**fresh, household appliances, digital products, and books**. JD.com first engaged in electrical appliances and digital products. As a result, household appliances and digital products are the two main business categories on JD.com with several product examples. Fresh and books are the two main new product categories developed in the past decade. Due to JD Logistics, many regions can enjoy the special logistics services of the next-day or even the same-day deliveries, which is very beneficial for the sale of fresh and books. Therefore, we select these four categories in level 1. For all categories except books, we exclusively obtain product titles. However, because of the unique nature of books, we collect titles as well as product descriptions. For the original dataset collected by us, we perform data cleaning to remove extraneous symbols and punctuation marks from the text. Furthermore, for each category in level 1, we randomly select 80% data as the training set and 20% data as the test set. An overview of our dataset is presented in Table 1. The number of categories in level 2 is close, all very few. However, number of categories in level 3 is large, and “household appliances” has more categories in level 3. In addition, the instances of “household appliances” are the most, while those of “books” are the least.

To train baseline algorithms using the code provided by Shimura et al. [42], we perform two key tasks: standardizing category names and constructing label trees to represent the hierarchical structure of our dataset. The categories in level 3 are renamed in the format “parent category in level 2@category in level 3”. For instance, if “novel” is a category in level 2 and “world classics” is a child category of “novel”, then “world classics” is renamed as “novel@world classics”. The hierarchical structure of our dataset is visually depicted in Figure 2, which illustrates the label tree. In addition, we provide examples from our dataset in Tables .5, .6, .7 and .8, showcasing one sample for each category in level 2.

Table 1: Summary of our dataset. The first row lists the four categories in level 1 of the dataset: fresh, household appliances, digital products, and books. We also report the number of categories in levels 2 and 3, the number of instances under each category in level 1, and the number of instances in the training and test sets under each category in level 1.

Categories in level 1	fresh	household appliances	digital products	books
# of categories in level 2	8	9	5	7
# of categories in level 3	80	172	50	85
Total sample size	132,175	580,226	279,116	19,933
sample size in the training set (80%)	105,740	464,180	223,292	15,946
sample size in the test set (20%)	26,435	116,046	55,824	3,987

4.2. Baseline Algorithms

We use three multilabel classifiers reported by Shimura et al. [42] as the baselines, i.e., the Flat-CNN, Hier-CNN, and HFT-CNN models.

The Flat-CNN model, as reported by Shimura et al. [42], is essentially a flat multilabel classifier. It treats the HTC problem as a flat multiclassification problem, ignoring the inherent hierarchical structure within the dataset. Our implementation incorporates a fastText layer for word embedding and utilizes a CNN architecture comprising three convolution layers, a max pooling layer, and a fully connected layer. To enhance model robustness, DropOut is applied within the fully connected layer.

The Hier-CNN model, as introduced by Shimura et al. [42], is a hierarchical multilabel classifier, that falls under the category of LCL approaches. The Hier-CNN approach involves the development and training of separate CNNs for each level within the hierarchy. Notably, the CNN architecture employed at each hierarchical level mirrors the structure used in the Flat-CNN model, which comprises three convolution layers, a max pooling layer, and a fully connected layer.

The HFT-CNN model proposed by Shimura et al. [42] is another hierarchical multilabel classifier, that belongs to the LCL approach. However, the primary difference between the Hier-CNN and HFT-CNN model lies in the incorporation of a fine-tuning mechanism. In the HFT-CNN model, an initial CNN is established, identical to the one employed in the Flat-CNN model, for the first hierarchical level. Subsequently, parameters derived from training of the upper levels are fine tuned at the lower levels.

Table 2 presents a comparative analysis of the three baseline algorithms along with our HFT-BERT model. We compare these four models from three aspects: whether they are flat, hierarchical, or fine-tuned. A model can only be either flat or hierarchical. The Flat-CNN is a flat model, while the other three models are hierarchical. The fine-tuning mechanism can be used for flat as well as hierarchy models. In our work, HFT-CNN as well as HFT-BERT contain fine-tuning mechanisms, whereas the other two models do not.

Table 2: Comparison of three baseline algorithms and based on HFT-BERT model over three aspects: flat, hierarchical, and fine-tuned models. The only flat model is the Flat-CNN, which is used as a contrast to the hierarchical models. Among the three hierarchical models, hier-CNN is used as a contrast to the fine-tuned models. HFT-CNN is used as a contrast to the BERT model.

Model	Flat	Hierarchical	Fine-tuned
Flat-CNN	✓	/	/
Hier-CNN	/	✓	/
HFT-CNN	/	✓	✓
HFT-BERT	/	✓	✓

4.3. Results

Experiments are conducted on the Ubuntu 18.04 operating system, which are built with the Pytorch framework, Chainer framework, CUDA 11.1 environment, and Python3.7 language. Hardware includes a GPU with NVIDIA Tesla P100-16GB and 64 GB of memory.

Before training the models, we first perform word segmentation on the preprocessed Chinese text. Our concern is the classification of categories in levels 2 and 3 under each category in level 1. As a result, we separate the dataset into four parts, each belonging to a category in level 1, i.e., fresh, household appliances, digital products, and books. The first 30 words are reserved for commodities under the categories in level 1, except books, and the first 200 words are reserved for books. For the baselines and our HFT-BERT model, fixed parameters include the embedding dimension of 300, and batch size of 128.

We train the baselines and our model on our dataset. The accuracy results of the four models at level 2 are shown in Table 3 and the accuracy results at level 3 are shown in Table 4. Flat-CNN achieves the best performance for categories fresh, household appliances, and digital products. This model ignores the inherent hierarchical structure within the dataset, predicting categories in levels 2 and 3 simultaneously. Therefore, it can learn more information than other models when predicting categories in level 2, making it easier to achieve the best performance. The accuracy of our model is close to those of the baselines for categories fresh, household, and digital products. Regarding books, the accuracy of our model is 3.5% higher than that of the best result among the three baselines and 5% higher than those of others.

Table 3: Accuracy results of the four models at level 2 under four categories at level 1. The best performances of the four models are shown in bold. Our model achieves similar performance for categories fresh and digital products and slightly inferior performance for household appliances. Our model achieves the best performance for books.

model	fresh	household appliances	digital products	books
Flat-CNN	**0.9841**	**0.9836**	**0.9854**	0.9143
Hier-CNN	0.9834	0.9827	0.9820	0.9095
HFT-CNN	0.9822	0.9828	0.9819	0.9008
HFT-BERT	0.9816	0.9724	0.9773	**0.9501**

To present the details of the classification results, we plot the confusion matrix of the categories in level 2 obtained by the HFT-BERT model (Figure 3). The provided confusion matrix depicts the performance of the HFT-

BERT model by presenting a square matrix with one row and one column for each category. The cells along the diagonal of the matrix represent the correctly classified instances for each category. Further, off-diagonal cells represent instances that are incorrectly classified, with each cell indicating the proportion of instances from one category that are mistakenly assigned to a different category. Through visual representation, we provide insights into the performance of HFT-BERT on longer short texts and potential areas of improvement.

The first picture on the upper left in Figure 3 displays the classification results for fresh. The numbers 0–7 correspond to “dairy products and cold drinks”, “fruit”, “aquatic products”, “pig beef and mutton”, “poultry and eggs”, “vegetables”, “fast food and prepared food”, and “pastry and baking”, respectively. The accuracy for “fast food and prepared food” is the lowest, which is less than 90%. The accuracies for all other categories reach 96% and above. Therefore, improving the classification accuracy of “fast food and prepared food” is the key to improve the overall classification performance of the model. The low accuracy could be due to two factors: on one hand, the number of instances in “fast food and prepared food” is the least among all categories, leading to an unsatisfactory model learning effect. On the other hand, “fast food and prepared food” is a very vague concept: for example, frozen steaks can be classified as either “pig beef and mutton” or “fast food”. The actual category depends on the category filled in by the merchants when encountering vague concepts, which is not standardized.

The second picture on the upper right displays the classification result for household appliances. The numbers 0–8 correspond to “personal health care”, “kitchen and bathroom electrics”, “small kitchen appliances”, “commercial appliances”, “large electrics”, “household appliance service”, “household appliance accessories”, “life electrics”, and “audio-video”, respectively. The accuracy for “commercial appliances” is the lowest, which is less than 90%. The accuracies for all other categories reach 95% and above. According to the misclassified cases of “commercial appliances”, we can infer that this is caused by an unclear concept definition. “Commercial appliances” may be incorrectly classified as “kitchen and bathroom electrics”, “small kitchen appliances”, or “large electrics”. For example, range hoods can be classified as “small kitchen appliances”, “kitchen and bathroom electrics”, “commercial appliances”, and “large electrics”. If the commodity name does not distinguish between home use and business use, it is logically correct to classify it into either category. It also depends on the merchant who subjectively determines the true classification, leading to an incorrect classification result.

The third picture on the lower left displays the classification results for digital products. Digital products are classified well, with accuracies above 97%.The numbers 0–4 correspond to “video entertainment”, “photography”, “digital accessories”, “intelligent equipment”, and “e-learning”, respectively. The last picture on the lower right displays the classification results for books. The numbers 0–6 correspond to “humanities and social societies”, “novel”, “life”, “science and technology”, “management”, “computer and Internet”, and “finance and investment”, respectively. Despite our model exhibiting the highest accuracy for books, there remains substantial scope for further refinement and optimization. Performance for “novel” is the best, achieving 100% accuracy. However, our model performs poorly when classifying instances of “humanities and social societies” and “life”. For instance, 12% of instances under “humanities and social societies” are incorrectly classified into “life”. This is because humanities, social sciences and life are closely related concepts. Further, 9% of the instances under “life” are incorrectly classified into “science and technology”, which may be because technology and life are always mentioned at the same time.

The accuracy of our model at level 3 is better on all four data sets. Our model shows an average increase of 4.5% on fresh, 2% on the household appliances, 1% on the digital products and 43% on books, offering advantages over the baselines, especially on books. Because of the excessive number of three-level categories, we do not show the confusion matrix of the classification. The performance of our model comparable with those of the other models under the first three categories (fresh, household appliances, and digital products) and superior performance for books show that our model can achieve performances comparable to those of CNN-based models on short texts and outperform CNN-based models on relatively longer short texts.

5. Conclusion

We introduce a novel approach for HTC using HFT-BERT. Based on the large-scale hierarchical dataset we collect, which comprises 1,011,450 products with titles and category names, our model achieves prediction performance comparable to those of existing methods. In particular, HFT-BERT demonstrates exceptional performance in categorizing books utilizing titles as well as product descriptions. Our model achieves short-text classification

Table 4: Accuracy results of the four models at level 3 under four categories at level 1. The best performances of the four models are shown in bold. Our model achieves the best performance on all datasets. In particular, our model shows improved performance compared with those of other models for books.

model	fresh	household appliances	digital products	books
Flat-CNN	0.9331	0.9480	0.9660	0.4929
Hier-CNN	0.9332	0.9520	0.9643	0.4744
HFT-CNN	0.9332	0.9482	0.9626	0.4809
HFT-BERT	**0.9828**	**0.9801**	**0.9814**	**0.9231**

performance comparable to those of other existing models for fresh, household appliances, and digital products. With respect to longer short texts, it demonstrates an excellent classification accuracy of 0.9231 for books. Therefore, the proposed model offers a promising approach for enhancing the efficiency and consistency of classification operations on Chinese e-commerce platforms. In the future, we will incorporate additional hierarchical levels into our model and compare its performance with those of state-of-the-art methods for HTC. In addition, we will expand our dataset further to include more categories.

Acknowledgments

The research of Rui Pan is supported by the National Natural Science Foundation of China (No. 72471254), the Program for Innovation Research, the Disciplinary Funds and the Emerging Interdisciplinary Project of Central University of Finance and Economics. Feifei Wang is supported by National Natural Science Foundation of China (No.72371241), the MOE Project of Key Research Institute of Humanities and Social Sciences (22JJD110001).

Conflict of interest

The authors declare no potential conflict of interests.

Table .5: Examples of our dataset. This table includes examples of each category at level 2 under fresh.

level 1	level 2	level 3	title
fresh	pastry and baking	pastry and baking@pizza	pizza pie semi-finished pizza ingredients 6/8/9 inches spaghetti pie artisanal pizza…
fresh	fruit	fruit@pitaya	domestic red pitaya (dragon fruit) 5 catty single fruit 300-400g healthy light meals…
fresh	aquatic products	aquatic products @sea cucumber	Beijing Tongrentang's quick hair sea cucumber 8g 3pcs bag dried aquatic products…
fresh	vegetables	vegetables @leafy vegetables	Mignon's Home home grown fresh oilseed cabbage 300g pot pot vegetables Beijing…
fresh	pig beef and mutton	pig beef and mutton@beef	Sanpin's four seasons Australian fatty beef rolls Angus beef chops grain-Fed M3 snowflake…
fresh	poultry and eggs	poultry and eggs@eggs	Nanyang agricultural specialties pavilion my hometown farm grain eggs mountain forest…
fresh	dairy products and cold drinks	dairy products and cold drinks @ice cream	Tianqiyipin's Ice Cream Ice Cream Chocolate Crunch Ice Cream Mint Yogurt…
fresh	fast food and prepared food	fast food and prepared food @meat products	BERETTA's Hungarian-style smoked salami cold chain delivery cuts…

Table .6: Examples of our dataset. This table includes examples of each category at level 2 under household appliances at level 1.

level 1	level 2	level 3	title
household appliances	small kitchen appliances	small kitchen appliances @decocting pot	Cuckoo’s a pot of a hundred ways to drink decocting pot automatic Chinese medicine…
household appliances	life electrics	life electrics @clothes driers	clothes dryer home clothes dryer double layer large capacity warm air speed drying small…
household appliances	personal health care	personal health care@hair dryer	Pinshile‘s Vertical electric hair dryer household hot and cold constant temperature quick…
household appliances	large electrics	large electrics @flat television	SHARP G70FL 4T-B70BHH5 70 inches 4K…
household appliances	kitchen and bathroom electrics	kitchen and bathroom electrics @dishwashers	TOSHIBA’s 13/14 sets of freestanding built-in dishwashers for home use…
household appliances	household appliances accessories	household appliances accessories@ television accessories	Xinshengtong/Skyworth’s LCD TV Remote Control Panel 50X3 55E…
household appliances	commercial electrics	commercial electrics @ice machine	Frestec’s ice machine for commercial use large square ice bar milk tea drinks…
household appliances	audio-video	audio-video @mini-speaker	PANDA’s 800 portable CD player tape recorder cassette player radio…
household appliances	household appliance service	household appliance service@household installation	gas cooker installation service

Table .7: Examples of our dataset. This table includes examples of each category at level 2 under digital products at level 1.

level 1	level 2	level 3	title
digital products	intelligent equipment	intelligent equipment @unmanned drone	HARWAR MEGA-V8 quadcopter octocopter UAV (customized version)
digital products	digital accessories	digital accessories @camera bag	Aibao‘s SLR camera 30/50/100/160L moisture proof box office home electronics…
digital products	photography	photography @camera lens	Shima’s 16mm F1.4 DC DN half-frame wide-angle lens Sony’s A6…
digital products	e-learning	e-learning @repeater	English listening mp3 Walkman mp4 portable learning machine repeater compact…
digital products	video entertainment	video entertainment@radio	for Tecsun R-909 radio elderly radio full band…

Table .8: Examples of dataset. This table includes examples of each category at level 2 under books at level 1.

level 1	level 2	level 3	title
books	computer and Internet	computer and Internet @artificial intelligence	“Machine Learning for Beginners” is a must-have book on machine learning, with no dizzying formulas but easy-to-understand analogies…
books	novel	novel@world classics	A selection of Maupassant’s novels Two covers shipped at random. Classic translation by Lee Yuk Min, featuring such masterpieces as “The Goat’s Ball”, “The Necklace”, and more!…
books	science and technology	science and technology @other industries	“Riveting and Welding Processing Quick Calculation” is based on what riveters and welders should know and be able to do as well as the basic techniques they should master, combining theory…
books	management	management @leadership	There are three parts to this book, the first part explains the concept of positive thought leadership, why the author uses positive thoughts to train the self, and others…
books	finance and investment	finance and investment @stock	The “New Stockbrokers Quick Start” starts with the quasi preparations and stock market terminology that new stockbrokers should do before entering the market, and then go-es on to talk about the…
books	life	life@cook and delicious food	The world of wine is complex, not only because there are so many different types of wines, but also because buying wine is a highly subjective endeavor…
books	humanities and social sciences	humanities and social sciences @psychology	The book focuses on three types of deep non-monotonic cognitive change - the creation of novelty, the adaptation of cognitive skills to changing environments, and the transformation of belief systems…

References

[1] S. Banerjee, C. Akkaya, F. Perez-Sorrosal, and K. Tsioutsiouliklis. Hierarchical transfer learning for multi-label text classification. In *inproceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6295–6300. Association for Computational Linguistics, 2019.

[2] L. Cai, Y. Song, T. Liu, and K. Zhang. A hybrid bert model that incorporates label semantics via adjustive attention for multi-label text classification. *IEEE Access*, 8:152183–152192, 2020.

[3] C.-M. Chen, H.-M. Lee, and C.-W. Hwang. A hierarchical neural network document classifier with linguistic feature selection. *Applied Intelligence*, 23:277–294, 2005.

[4] H. Chen, Q. Ma, Z. Lin, and J. Yan. Hierarchy-aware label semantics matching network for hierarchical text classification. In *inproceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4370–4379. Association for Computational Linguistics, 2021.

[5] H. Chen, J. Zhao, and D. Yin. Fine-grained product categorization in e-commerce. In *inproceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2349–2352. Association for Computing Machinery, 2019.

[6] A. Clare and R. D. King. Predicting gene function in saccharomyces cerevisiae. *Bioinformatics*, 19(suppl_2):ii42–ii49, 2003.

[7] S. D'Alessio, K. Murray, R. Schiaffino, and A. Kershenbaum. The effect of using hierarchical classifiers in text categorization. In *Content-Based Multimedia Information Access - Volume 1*, page 302–313. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 2000.

[8] O. Dekel, J. Keshet, and Y. Singer. Large margin hierarchical classification. In *inproceedings of the Twenty-First International Conference on Machine Learning*, page 27. Association for Computing Machinery, 2004.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[10] S. Dhote, C. Vichoray, R. Pais, S. Baskar, and P. Mohamed Shakeel. Hybrid geometric sampling and adaboost based deep learning approach for data imbalance in e-commerce. *Electronic Commerce Research*, 20(2):259–274, 2020.

[11] R. Duwairi and R. Al-Zubaidi. A hierarchical K-NN classifier for textual data. *The International Arab Journal of Information Technology (IAJIT)*, 8(3):251–259, 2011.

[12] A. Esuli, T. Fagni, and F. Sebastiani. TreeBoost. MH: A boosting algorithm for multi-label hierarchical text categorization. In *String Processing and Information Retrieval*, pages 13–24. Springer Berlin Heidelberg, 2006.

[13] Y. Fang, C. Ma, D. Wu, and T. Zhang. Commodity classification based on feature enhancement. In *International Conference on Electronic Information Engineering and Data Processing (EIEDP 2023)*, volume 12700, pages 433–438. SPIE, 2023.

[14] F. Gargiulo, S. Silvestri, M. Ciampi, and G. De Pietro. Deep neural network for hierarchical extreme multi-label text classification. *Applied Soft Computing*, 79:125–138, 2019.

[15] E. Gaussier, C. Goutte, K. Popat, and F. Chen. A hierarchical model for clustering and categorising documents. In *inproceedings of the 24th BCS-IRSG European Colloquium on IR Research: Advances in Information Retrieval*, pages 229–247. Springer-Verlag, 2002.

[16] J. Gong, Z. Teng, Q. Teng, H. Zhang, L. Du, S. Chen, M. Z. A. Bhuiyan, J. Li, M. Liu, and H. Ma. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification. *IEEE Access*, 8:30885–30896, 2020.

[17] I. Hasson, S. Novgorodov, G. Fuchs, and Y. Acriche. Category recognition in e-commerce using sequence-to-sequence hierarchical classification. In *inproceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 902–905. Association for Computing Machinery, 2021.

[18] L. He, Y. Jia, Z. Ding, and W. Han. Hierarchical classification with a topic taxonomy via LDA. *International Journal of Machine Learning and Cybernetics*, 5(4):491–497, 2014.

[19] W. Huang, E. Chen, Q. Liu, Y. Chen, Z. Huang, Y. Liu, Z. Zhao, D. Zhang, and S. Wang. Hierarchical multi-label text classification: An attention-based recurrent network approach. In *inproceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1051–1060. Association for Computing Machinery, 2019.

[20] JD.com. Helpcenter, 2023. `https://helpcenter.jd.com`, Last accessed on 2023-11-22.

[21] S. Khan and A. R. Baig. Ant colony optimization based hierarchical multi-label classification algorithm. *Applied Soft Computing*, 55:462–479, 2017.

[22] S. Khan, A. R. Baig, and W. Shahzad. A novel ant colony optimization based single path hierarchical classification algorithm for predicting gene ontology. *Applied Soft Computing*, 16:34–49, 2014.

[23] S. Kiritchenko, S. Matwin, R. Nock, and A. F. Famili. Learning and evaluation in the presence of class hierarchies: Application to text categorization. In *Conference of the Canadian society for computational studies of intelligence*, pages 395–406. Springer, 2006.

[24] D. Koller and M. Sahami. Hierarchically classifying documents using very few words. In *inproceedings of the Fourteenth International Conference on Machine Learning*, pages 170–178. Morgan Kaufmann Publishers Inc., 1997.

[25] J. Kong, J. Wang, and X. Zhang. Hierarchical bert with an adaptive fine-tuning strategy for document classification. *Knowledge-Based Systems*, 238:107872, 2022.

[26] K. Kowsari, D. E. Brown, M. Heidarysafa, K. J. Meimandi, M. S. Gerber, and L. E. Barnes. Hdltex: Hierarchical deep learning for text classification. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 364–371. IEEE, 2017.

[27] Z. Kozareva. Everyone likes shopping! multi-class product categorization for e-commerce. In *inproceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1329–1333. Association for Computational Linguistics, 2015.

[28] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[29] H. Lee and Y. Yoon. Engineering doc2vec for automatic classification of product descriptions on o2o applications. *Electronic Commerce Research*, 18:433–456, 2018.

[30] J. Leskovec and A. Krevl. SNAP datasets: Stanford large network dataset collection, 2014.

[31] D. D. Lewis, Y. Yang, T. Russell-Rose, and F. Li. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5:361–397, 2004.

[32] G. Li and N. Li. Customs classification for cross-border e-commerce based on text-image adaptive convolutional neural network. *Electronic Commerce Research*, 19(4):779–800, 2019.

[33] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang. Deep learning for extreme multi-label text classification. In *inproceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 115–124. Association for Computing Machinery, 2017.

[34] J. Liu, C. Xia, H. Yan, Z. Xie, and J. Sun. Hierarchical comprehensive context modeling for chinese text classification. *IEEE Access*, 7:154546–154559, 2019.

[35] K. Ma, Z. Huang, X. Deng, J. Guo, and W. Qiu. LED: Label correlation enhanced decoder for multi-label text classification. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[36] Y. Ma, X. Liu, L. Zhao, Y. Liang, P. Zhang, and B. Jin. Hybrid embedding-based text representation for hierarchical multi-label text classification. *Expert Systems with Applications*, 187:115905, 2022.

[37] R. Moskovitch, S. Cohen-Kashi, U. Dror, I. Levy, A. Maimon, and Y. Shahar. Multiple hierarchical classification of free-text clinical guidelines. *Artificial Intelligence in Medicine*, 37(3):177–190, 2006.

[38] C. D. Nguyen, T. A. Dung, and T. H. Cao. Text classification for DAG-structured categories. In *Advances in Knowledge Discovery and Data Mining*, pages 290–300. Springer Berlin Heidelberg, 2005.

[39] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, and Q. Yang. Large-scale hierarchical text classification with recursively regularized deep graph-CNN. In *inproceedings of the 2018 World Wide Web Conference*, pages 1063–1072. International World Wide Web Conferences Steering Committee, 2018.

[40] A. Secker, M. Davies, A. Freitas, E. Clark, J. Timmis, and D. Flower. Hierarchical classification of g-protein-coupled receptors with data-driven selection of attributes and classifiers. *International Journal of Data Mining and Bioinformatics*, 4(2):191–210, 2010.

[41] A. D. Secker, M. N. Davies, A. A. Freitas, J. Timmis, M. Mendao, and D. R. Flower. An experimental comparison of classification algorithms for hierarchical prediction of protein function. *Expert Update (Magazine of the British Computer Society's Specialist Group on AI)*, 9(3):17–22, 2007.

[42] K. Shimura, J. Li, and F. Fukumoto. HFT-CNN: Learning hierarchical category structure for multi-label short text categorization. In *inproceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 811–816. Association for Computational Linguistics, 2018.

[43] C. Silla and A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22:31–72, 2011.

[44] C. N. Silla Jr. and A. A. Freitas. A global-model naive bayes approach to the hierarchical prediction of protein functions. In *2009 Ninth IEEE International Conference on Data Mining*, pages 992–997. IEEE, 2009.

[45] K. Sinha, Y. Dong, J. C. K. Cheung, and D. Ruths. A hierarchical neural attention-based text classifier. In *inproceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 817–823. Association for Computing Machinery, 2018.

[46] J. Song, F. Wang, and Y. Yang. Peer-label assisted hierarchical text classification. In *inproceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3747–3758. Association for Computational Linguistics, 2023.

[47] A. Sun and E.-P. Lim. Hierarchical text classification and evaluation. In *inproceedings 2001 IEEE International Conference on Data Mining*, pages 521–528. IEEE Computer Society, 2001.

[48] T. Thaminkaew, P. Lertvittayakumjorn, and P. Vateekul. Prompt-based label-aware framework for few-shot multi-label text classification. *IEEE Access*, 12:28310–28322, 2024.

[49] P. Vateekul, M. Kubat, and K. Sarinnapakorn. Hierarchical multi-label classification with SVMs: A case study in gene function prediction. *Intelligent Data Analysis*, 18(4):717–738, 2014.

[50] C. Wang, H. Jiang, T. Chen, J. Liu, M. Wang, S. Jiang, Z. Li, and Y. Xiao. Entity understanding with hierarchical graph learning for enhanced text classification. *Knowledge-Based Systems*, 244:108576, 2022.

[51] F. Wu, J. Zhang, and V. Honavar. Learning classifiers using hierarchically structured class taxonomies. In *inproceedings of the 6th International Conference on Abstraction, Reformulation and Approximation*, pages 313–320. Springer-Verlag, 2005.

[52] Z. Yang and G. Liu. Hierarchical sequence-to-sequence model for multi-label text classification. *IEEE Access*, 7:153012–153020, 2019.

[53] J. Zhang, Y. Li, F. Shen, C. Xia, H. Tan, and Y. He. Hierarchy-aware and label balanced model for hierarchical text classification. *Knowledge-Based Systems*, 300:112153, 2024.

[54] R. Zhang, H. Lee, and D. R. Radev. Dependency sensitive convolutional neural networks for modeling sentences and documents. In *inproceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1512–1521. Association for Computational Linguistics, 2016.

[55] W. Zhang, F. Liu, Z. Zhang, S. Liu, and Q. Huang. Commodity text classification based e-commerce category and attribute mining. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 105–108. IEEE, 2020.

[56] X. Zhang, J. Xu, C. Soh, and L. Chen. LA-HCN: label-based attention for hierarchical multi-label text classification neural network. *Expert Systems with Applications*, 187:115922, 2022.

[57] J. Zhou, C. Ma, D. Long, G. Xu, N. Ding, H. Zhang, P. Xie, and G. Liu. Hierarchy-aware global model for hierarchical text classification. In *inproceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1106–1117. Association for Computational Linguistics, 2020.

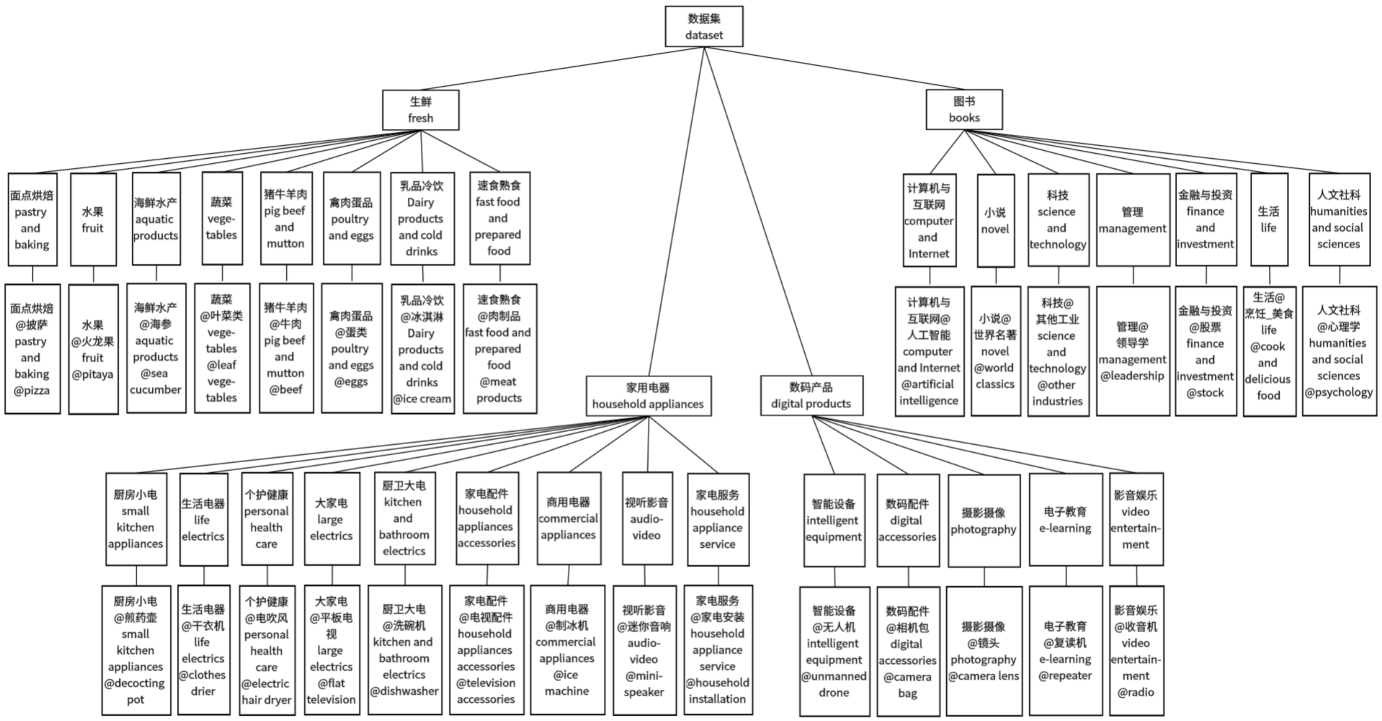
数据集 dataset
生鲜 fresh
家用电器 household appliances
数码产品 digital products
图书 books
面点烘焙 pastry and baking
水果 fruit
海鲜水产 aquatic products
蔬菜 vege-tables
猪牛羊肉 pig beef and mutton
禽肉蛋品 poultry and eggs
乳品冷饮 Dairy products and cold drinks
速食熟食 fast food and prepared food
面点烘焙 @披萨 pastry and baking @pizza
水果 @火龙果 fruit @pitaya
海鲜水产 @海参 aquatic products @sea cucumber
蔬菜 @叶菜类 vege-tables @leaf vege-tables
猪牛羊肉 @牛肉 pig beef and mutton @beef
禽肉蛋品 @蛋类 poultry and eggs @eggs
乳品冷饮 @冰淇淋 Dairy products and cold drinks @ice cream
速食熟食 @肉制品 fast food and prepared food @meat products
计算机与互联网 computer and Internet
小说 novel
科技 science and technology
管理 management
金融与投资 finance and investment
生活 life
人文社科 humanities and social sciences
计算机与互联网@人工智能 computer and Internet @artificial intelligence
小说@世界名著 novel @world classics
科技@其他工业 science and technology @other industries
管理@领导学 management @leadership
金融与投资 @股票 finance and investment @stock
生活@烹饪_美食 life @cook and delicious food
人文社科 @心理学 humanities and social sciences @psychology
厨房小电 small kitchen appliances
生活电器 life electrics
个护健康 personal health care
大家电 large electrics
厨卫大电 kitchen and bathroom electrics
家电配件 household appliances accessories
商用电器 commercial appliances
视听影音 audio-video
家电服务 household appliance service
厨房小电 @煎药壶 small kitchen appliances @decocting pot
生活电器 @干衣机 life electrics @clothes drier
个护健康 @电吹风 personal health care @electric hair dryer
大家电 @平板电视 large electrics @flat television
厨卫大电 @洗碗机 kitchen and bathroom electrics @dishwasher
家电配件 @电视配件 household appliances accessories @television accessories
商用电器 @制冰机 commercial appliances @ice machine
视听影音 @迷你音响 audio-video @mini-speaker
家电服务 @家电安装 household appliance service @household installation
智能设备 intelligent equipment
数码配件 digital accessories
摄影摄像 photography
电子教育 e-learning
影音娱乐 video entertain-ment
智能设备 @无人机 intelligent equipment @unmanned drone
数码配件 @相机包 digital accessories @camera bag
摄影摄像 @镜头 photography @camera lens
电子教育 @复读机 e-learning @repeater
影音娱乐 @收音机 video entertain-ment @radio

Figure 2: Label tree of the dataset. Categories in level 1 of the dataset include fresh, household appliances, digital products, and books. For each category in level 1, all subcategories in level 2 are shown. For each category in level 2, only one of the subcategories in level 3 is shown.

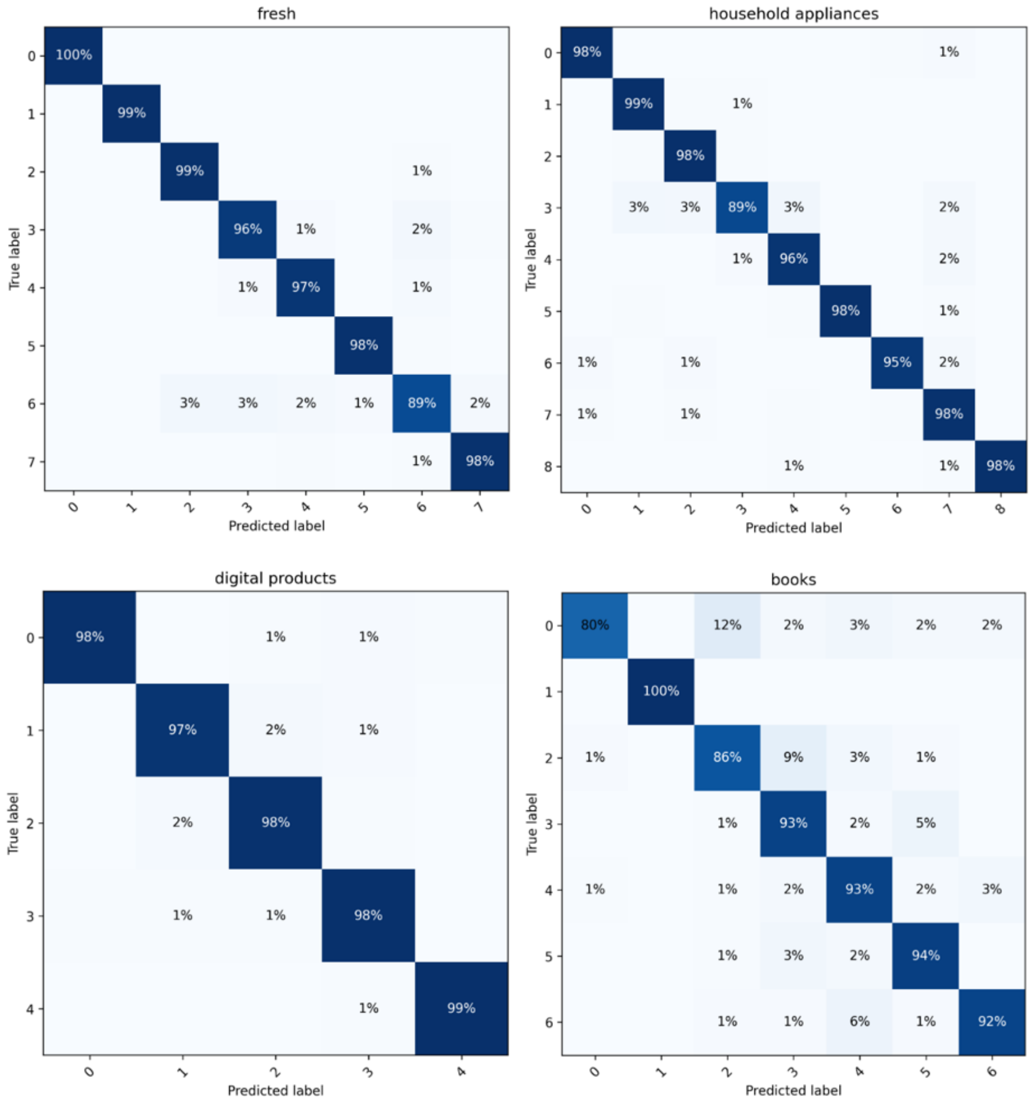

Figure 3: Confusion matrix of categories in level 2 obtained from the HFT-BERT model. The values at the diagonal positions represent the classification accuracy of each category in level 2. The values at the off-diagonal positions represent the misclassification rate of each category in level 2 classified into other categories.