Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Università Ca'Foscari Venezia

Hierarchical Text Classification: A Review Of Current Research

Zangari Alessandro, Marcuzzo Matteo, Schiavinato Michele, Giudice Lorenzo, Albarelli Andrea, Gasparetto Andrea

iNEST Spoke 6 “Tourism, Culture and Creative Industries”
RT: 1 Sub. RT: 3

Abstract

It is often the case that collections of documents are annotated with hierarchically structured concepts. However, the benefits of this structure are rarely considered by commonly used classification techniques. Conversely, Hierarchical Text Classification methods are devised to boost classification performance by taking advantage of the labels’ organization. This work aims to deliver an updated overview of current research in this domain. We define and frame the task within the broader text classification area, examining important shared concepts such as text representation. Then, we dive into details regarding the specific task, providing a high-level description of its traditional approaches. We then summarise recently proposed methods, highlighting their main contributions. We additionally provide statistics for the most adopted datasets and describe the benefits of using evaluation metrics tailored to hierarchical settings. Finally, a selection of recent proposals is benchmarked against non-hierarchical baselines on five domain-specific datasets.

Text

Hierarchical Text Classification (HTC) is a specialized area of Text Classification (TC) that has gained increasing importance over the years. It focuses on organizing and classifying documents into a hierarchical structure of concepts or categories. This field is particularly relevant in contexts where information is organized hierarchically, such as academic literature, legal documents, or website categorization. HTC techniques offer several advantages over standard flat classification techniques. They use the inherent structure of label hierarchies, leading to improved classification performance and a more nuanced understanding of textual data. One of the key challenges in HTC is the design of an effective hierarchy of categories. This involves identifying the right concepts and categories and defining their relationships. HTC techniques can be broadly classified into *top-down* and *bottom-up* categories (Yu et al., 2022; Zhang et al., 2022). *Top-down* approaches start with a high-level category and recursively divide it into subcategories until the desired level of granularity is achieved. *Bottom-up* approaches, on the other hand, start with a set of documents and cluster them into categories based on similarities in their content.

HTC has several real-world applications. For academic literature, it can classify research papers into various disciplines and sub-disciplines (Huang et al., 2019; Sharma et al., 2022). In legal documents, it can be used to classify case law into various legal domains and sub-domains (Caled et al., 2019). For website categorization, it can group web pages into topics and subtopics (Zhao et al., 2021). HTC offers several advantages over standard flat classification techniques and has several real-world applications (Sun et al., 2003; Sun and Lim, 2001a). With the increasing availability of textual data and the need for automated classification, the importance of HTC is only set to grow in the future.

Related literature

In HTC, labels are structured hierarchically, typically represented as a tree or a directed acyclic graph (DAG) (Peng et al., 2018; Wang et al., 2022). Each node in this structure represents a category, and its position reflects the conceptual relationship with other categories. Considering its content and hierarchical relevance, this arrangement allows for a more granular and contextual text classification. The complexity of real-world datasets makes HTC a valuable approach for systematically managing and interpreting large volumes of text data. The primary aim of this field is to enhance the accuracy and efficiency of text classification by leveraging the relationships between hierarchical labels (Ceci and Malerba, 2007; Sun et al., 2004, p. 200).

HTC finds significant application in domains with large and complex category systems, such as digital libraries and e-commerce platforms (Li et al., 2022). In these areas, effective classification requires understanding not only the content of the documents but also their position within the broader category structure. The hierarchical nature of

these domains means that HTC is not just a methodological choice but a necessity for accurate and efficient information retrieval and organization.

The evolution of HTC is marked by a transition from early methods, which focused on addressing the limitations of flat classifiers, to more contemporary approaches that exploit advanced machine learning and deep learning techniques. Seminal works in the field, such as those by (Koller and Sahami, 1997), identified the shortcomings of traditional classifiers when dealing with hierarchical data structures. These foundational works paved the way for more sophisticated methods considering hierarchical relationships in classification tasks. Subsequent research has expanded on these ideas, proposing various approaches to effectively integrate hierarchical information into classification models (Marcuzzo et al., 2022; Silla and Freitas, 2011).

Recent reviews in the field, such as those by (Silla and Freitas, 2011; Stein et al., 2019), offer a comprehensive overview of hierarchical classification, comparing different methodologies and their effectiveness in various applications. These reviews critically evaluate both traditional and neural model-based approaches to HTC. They highlight the advancements in the field, demonstrating how modern techniques have overcome earlier models' limitations and provided more accurate and efficient classification systems.

This study contributes to the field by analyzing recent trends in HTC research, focusing on publications from 2019 to 2022. We collected papers using targeted keywords related to HTC from academic databases, providing a well-rounded view of the latest developments in this area. The main contributions include an extensive review of current HTC research, an exploration of NLP background with a focus on text representation and neural architectures, an analysis of common approaches to HTC, and evaluation measures. It also includes a summary of recent proposals for HTC, benchmarking these against non-hierarchical baselines. This comprehensive approach to reviewing HTC literature highlights the current state of the art and underscores this field's dynamic nature. As new challenges and datasets emerge, HTC continues to evolve, adapting to the changing landscape of TC. Integrating advanced neural network architectures and exploring novel representation techniques indicate the field's ongoing commitment to improving classification accuracy and efficiency. Furthermore, applying HTC in diverse domains underscores its versatility and the growing recognition of its importance in various sectors.

NLP Background and Methodologies

HTC is a significant study area in the dynamic NLP field, intertwining advanced text representation techniques with innovative neural architectures. This exploration begins by tracing the evolution of text representation, a journey marked by pivotal developments that have profoundly influenced HTC.

The progression from basic statistical methods like word counts to sophisticated text representation techniques underscores the advancements in NLP. Early milestones

include the development of Term Frequency-Inverse Document Frequency (TF-IDF), a method for weighted word counts (Salton and McGill, 1986), which laid the groundwork for more complex models. The introduction of word embeddings, such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), marked a significant leap forward. These embeddings represent words in a vector space, capturing their contextual meanings beyond mere occurrence.

The advent of contextualized language models, particularly those utilizing Transformer architectures (Vaswani et al., 2017), represents a quantum leap in text representation. These models dynamically adapt embeddings based on the surrounding context, leading to more nuanced and accurate text classification. This advancement has been crucial in enhancing the quality of text representation, a cornerstone in developing effective HTC methods. In parallel, the evolution of neural architectures has been instrumental in NLP and HTC. Recurrent Neural Networks (RNNs) have been foundational for sequential data processing, with subsequent enhancements like Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Units (GRU) (Cho et al., 2014) addressing their initial limitations. Though primarily celebrated in computer vision, convolutional Neural Networks (CNNs) have found a niche in NLP for extracting features from word embeddings (Kim, 2014; Lea et al., 2017; Yan et al., 2020). A groundbreaking innovation in NLP is the Transformer network, renowned for its self-attention layers that adeptly learn dependencies between tokens in text. This architecture has been the basis for models like BERT (Devlin et al., 2019) and GPT (Radford et al., 2018), which have redefined benchmarks in text classification performance.

HTC, as a specialized field within NLP, leverages these advancements in text representation and neural architectures. It employs unique approaches that utilize label hierarchies to enhance classification performance, distinguishing it from standard text classifiers. These approaches are broadly categorized into *local* and *global* methods (Yu et al., 2022). Local methods break down the hierarchical structure into manageable segments, treating each as a distinct classification problem. In contrast, global methods use a single classifier to consider the entire hierarchy, often yielding more efficient and accurate results due to their holistic view of interrelations and dependencies within the hierarchy.

HTC is not without its challenges, particularly in the nuanced decision-making required in hierarchical structures, such as determining the necessary depth of classification (Silla and Freitas, 2011; Zhou et al., 2020). Strategies to address these challenges include adaptive thresholds in local classifiers and specific modifications in global classifiers (Cerri et al., 2011; Mao et al., 2019; Punera and Ghosh, 2008).

Recent years have seen a surge in novel HTC methodologies driven by the continuous evolution of machine learning and NLP. The adoption of neural network architectures in HTC has been transformative. CNNs, RNNs, LSTM, GRU, and Transformer-based models have each contributed uniquely to understanding and processing hierarchical

text data. Innovative approaches in recent HTC research include using graph neural networks (GNNs) to capture complex label relationships and capsule networks to understand hierarchical data structures. Ensemble methods, which combine the strengths of multiple classifiers, have also been employed to improve classification accuracy and robustness. Furthermore, developing domain-specific HTC models, such as those tailored for legal or medical text classification, highlights the field's adaptability. These models, fine-tuned to the unique characteristics of their respective domains, leverage domain-specific knowledge to achieve enhanced performance (Gasparetto et al., 2022; Kowsari et al., 2019; Zhang and Wallace, 2017).

Benchmarking these models on standard datasets and employing appropriate evaluation metrics cannot be overstated. This process is vital for assessing the effectiveness of various HTC methodologies, identifying their strengths, and pinpointing areas for improvement.

In conclusion, HTC's journey reflects the broader evolution of NLP, continually pushing the boundaries of what is possible in understanding and classifying complex hierarchical text data. The exploration of HTC has become increasingly sophisticated, particularly in experimental approaches. This comprehensive analysis delves into the nuances of benchmarking recent HTC methods, testing novel proposals, and scrutinizing the methodologies and results of these experiments. Each facet of this exploration contributes significantly to our understanding of HTC's capabilities and limitations in various contexts.

Evaluating HTC

HTC inherently involves multiclass or multilabel issues. Many researchers opt to use standard evaluation metrics commonly adopted in classification scenarios. However, several authors argue that these measures may be inappropriate due to the hierarchical structure of the data (Kiritchenko et al., 2006; Kosmopoulos et al., 2015; Stein et al., 2019; Sun and Lim, 2001a). The argument is based on mistake severity, suggesting that a model making “better mistakes” should be preferred (Vaswani et al., 2022). This concept is rooted in two considerations. Firstly, predicting a label structurally close to the ground truth should be less penalizing than predicting a distant one. Secondly, errors at the upper levels of the hierarchy are inherently worse.

Regarding datasets used in HTC, the literature is spread across a wide variety of datasets, often derived from the same source but used in different ways. Some of the most prominent sources of raw data include DBPedia and Wikipedia. However, there is a lack of established benchmarks in HTC, resulting in methods being scattered over a wide range of incomparable datasets. Some datasets provide pre-defined splits, like the popular Reuters Corpus-V1 (RCV1) and Web of Science (WOS). However, caution is needed when comparing results, as some methods may use larger training splits, which could affect comparisons. We used five diverse datasets in our experimental setting to test recent HTC methods across domains. The first dataset is the *WOS*. This dataset, introduced by (Kowsari et al., 2018), is widely used in HTC research. It

comprises scientific literature data. The second dataset, *Blurb Genre Collection*, provided by (Aly et al., 2019), comes with predefined training and test splits. It contains book blurbs from various genres. The third dataset, *RCV1*, introduced by (Lewis et al., 2004), is a collection of Reuters News articles collected between August 20, 1996, and August 19, 1997. The dataset is manually categorized and contains over 800,000 international newswire stories in English. The fourth dataset is *Linux Bugs*. This dataset, introduced by (Lyubinets et al., 2018), comprises bug reports scraped from the Linux kernel bug tracker. The documents are support tickets classified in importance, related product, and specific component. Finally, the fifth dataset is *Amazon Reviews* (Ni et al., 2019). This dataset is derived from the Amazon corpus and contains user reviews. These datasets represent five application domains of HTC methods: books, scientific literature, news, IT tickets, and reviews. Such diversity allows us to test the methods across a wide spectrum of data. A specific preprocessing procedure for each method is developed and discussed in the paper.

In the experimental part of the study, various methods are tested for performance on the five above mentioned datasets. The selection was refined to include methods used as baselines in previous works. However, challenges such as missing dependencies, outdated libraries, and lack of instructions hindered the use of many existing implementations. Despite these obstacles, many methods were successfully tested. The methods tested included MATCH (Zhang et al., 2021), HiAGM (Zhou et al., 2020), BERT (Devlin et al., 2019), XML-CNN (Liu et al., 2017), and one SVM. MATCH uses a normalization mechanism for the labels and Word2Vec embeddings hierarchy, which involves text preprocessing and the Adam optimizer. The vastly adopted pre-trained "bert-base-cased" model from the HuggingFace library was also investigated. The aim was not to provide the best results for each method but to compare across datasets and methods and assess the ease of applicability of these methods on diverse datasets. Supplementary material was provided for further details on the validation procedure.

Challenges

In HTC, researchers grapple with several critical challenges that are pivotal to advancing this field. These challenges range from the intricacies of leaf node prediction to the complexities of handling imbalanced datasets, integrating domain-specific knowledge, ensuring scalability and computational efficiency, and enhancing model interpretability. The field of HTC faces a range of complex challenges that require innovative solutions and interdisciplinary approaches. Addressing these challenges is crucial for advancing intelligent text classification systems that are accurate, efficient, and trustworthy.

One of the foremost challenges is the *non-mandatory leaf node prediction* (Freitas and de Carvalho, 2007; Sun and Lim, 2001b). This issue becomes particularly significant in hierarchical structures where a detailed classification to the most specific level may not always be necessary or relevant. In such scenarios, it is essential to have intelligent

systems capable of determining the appropriate depth of classification based on the context. This challenge is not just about accuracy but also about the relevance of the classification in a given context (Silla and Freitas, 2011). Researchers are exploring various methodologies and strategies, such as adaptive algorithms and context-aware systems, to address this challenge effectively.

Another critical issue in HTC is *handling imbalanced datasets*, a common occurrence where some categories are underrepresented. This imbalance can significantly skew the learning process of HTC models, leading to poor performance in minority classes. Strategies like data augmentation, specialized loss functions, and re-sampling techniques are being explored to combat this. These methods aim to create a more balanced dataset, thereby improving the learning process and accuracy of the models (Vaswani et al., 2022; Zhou et al., 2020).

Integrating domain-specific knowledge into HTC models is also a crucial area of focus. This approach involves leveraging expert knowledge and domain-specific features to improve classification accuracy, particularly in specialized fields like legal or medical text classification. Integrating such knowledge can significantly enhance the performance of HTC models by providing them with a richer context and more relevant features for classification (Mencía and Fürnkranz, 2008; Rios and Kavuluru, 2018).

The *scalability and computational efficiency* challenge is particularly pertinent in handling large-scale datasets and complex hierarchies. As the volume of data grows, the need for more efficient algorithms and models that can process large volumes of data without compromising classification accuracy becomes increasingly important. Researchers are exploring various approaches, including advanced machine learning techniques and optimized algorithms, to improve scalability and efficiency in HTC models (Aljedani et al., 2021; Aly et al., 2019).

Lastly, the importance of *model interpretability*, especially deep learning-based HTC models, cannot be overstated. In sensitive applications, models must be accurate and transparent in their decision-making processes. Developing interpretable models is essential for their acceptance and trust among users. This involves creating models that provide clear insights into how and why they arrived at a particular classification decision (Madsen et al., 2022; Saranya and Subhashini, 2023)

Conclusion

This comprehensive review summarizes the significant findings, emphasizing the remarkable advancements in the HTC field. The review, a pivotal contribution to the literature, underscores the shift towards neural network-based approaches, a trend that has revolutionized the HTC landscape. Developing sophisticated models, particularly those that effectively leverage hierarchical information, marks a significant milestone in the evolution of text classification methodologies.

In deep learning for HTC, this paper highlights the instrumental role played by advanced techniques such as Transformer-based models, CNNs, RNNs, GNNs, and Capsule Networks. These methods have been lauded for their effectiveness in handling the complexities inherent in hierarchical data. Their ability to set new benchmarks in HTC is a testament to the rapid progress in this field. For instance, (Vaswani et al., 2017), in their groundbreaking paper on Transformer models, have laid the foundation for significant advancements in natural language processing, directly impacting HTC's progress.

The importance of benchmarking in HTC cannot be overstated. This review stresses the need for continuous and rigorous benchmarking of HTC models using diverse and challenging datasets. Such a process is crucial for assessing the effectiveness of various approaches, identifying their strengths and weaknesses, and guiding future research. Benchmarking, as highlighted by (Zhang et al., 2015) in their analysis of CNNs for sentence classification, provides invaluable insights into the performance and applicability of different models.

Looking ahead, the paper outlines several promising areas for future research. These include the development of models capable of handling non-mandatory leaf node prediction, which remains a challenging aspect of HTC. Additionally, there is a need for algorithms that are robust to imbalanced datasets, a common issue in real-world applications. Integrating domain-specific knowledge into HTC models is another area ripe for exploration, as it can significantly enhance the accuracy and relevance of classification. Furthermore, improvements in scalability and efficiency are essential for handling the ever-increasing volume of textual data. Lastly, enhancements in model interpretability are crucial for gaining insights into these complex models' decision-making processes, as (Guidotti et al., 2018) emphasised in their comprehensive review of interpretability in machine learning.

We conclude with an optimistic outlook on the future of HTC. While acknowledging the challenges, we highlight the rapid advancements in the field and the potential for further development. The authors hope that ongoing research will address the current limitations and explore new applications and methodologies in HTC. This optimism is rooted in the belief that the field of HTC, as demonstrated by its recent progress, holds immense potential for innovation and advancement, paving the way for more sophisticated and effective text classification solutions.

In summary, this review not only provides a thorough overview of the current state of HTC but also charts a course for future research, encouraging continued exploration and innovation in this dynamic and ever-evolving field.

References

Aljedani, N., Alotaibi, R., Taileb, M., 2021. HMATC: Hierarchical multi-label Arabic text classification model using machine learning. Egypt. Inform. J. 22, 225–237. https://doi.org/10.1016/j.eij.2020.08.004

Aly, R., Remus, S., Biemann, C., 2019. Hierarchical Multi-label Classification of Text with Capsule Networks, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Association for Computational Linguistics, Florence, Italy, pp. 323–330. https://doi.org/10.18653/v1/P19-2045

Caled, D., Won, M., Martins, B., Silva, M.J., 2019. A Hierarchical Label Network for Multi-label EuroVoc Classification of Legislative Contents, in: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (Eds.), Digital Libraries for Open Knowledge. Springer International Publishing, Cham, pp. 238–252. https://doi.org/10.1007/978-3-030-30760-8_21

Ceci, M., Malerba, D., 2007. Classifying web documents in a hierarchy of categories: a comprehensive study. J. Intell. Inf. Syst. 28, 37–78. https://doi.org/10.1007/s10844-006-0003-2

Cerri, R., Barros, R.C., de Carvalho, A.C.P.L.F., 2011. Hierarchical multi-label classification for protein function prediction: A local approach based on neural networks, in: 2011 11th International Conference on Intelligent Systems Design and Applications. pp. 337–343. https://doi.org/10.1109/ISDA.2011.6121678

Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches, in: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Association for Computational Linguistics, Doha, Qatar, pp. 103–111. https://doi.org/10.3115/v1/W14-4012

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. 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). Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423

Freitas, A., de Carvalho, A., 2007. A Tutorial on Hierarchical Classification with Applications in Bioinformatics. Res. Trends Data Min. Technol. Appl. https://doi.org/10.4018/978-1-59904-271-8.ch007

Gasparetto, A., Marcuzzo, M., Zangari, A., Albarelli, A., 2022. A Survey on Text Classification Algorithms: From Text to Predictions. Information 13, 83. https://doi.org/10.3390/info13020083

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A Survey of Methods for Explaining Black Box Models. ACM Comput. Surv. 51, 93:1-93:42. https://doi.org/10.1145/3236009

Hochreiter, S., Schmidhuber, J., 1997. Long Short-term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, W., Chen, E., Liu, Q., Chen, Y., Huang, Z., Liu, Y., Zhao, Z., Zhang, D., Wang, S., 2019. Hierarchical Multi-label Text Classification: An Attention-based Recurrent Network Approach, in: Zhu, W., Tao, D., Cheng, X., Cui, P., Rundensteiner, E.A., Carmel, D., He, Q., Yu, J.X. (Eds.), Proceedings of the 28th ACM International Conference on Information and Knowledge

Management, CIKM 2019, Beijing, China, November 3-7, 2019. ACM, pp. 1051–1060. https://doi.org/10.1145/3357384.3357885

Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, pp. 1746–1751. https://doi.org/10.3115/v1/D14-1181

Kiritchenko, S., Matwin, S., Nock, R., Famili, A.F., 2006. Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization, in: Lamontagne, L., Marchand, M. (Eds.), Advances in Artificial Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 395–406.

Koller, D., Sahami, M., 1997. Hierarchically Classifying Documents Using Very Few Words, in: Proceedings of the Fourteenth International Conference on Machine Learning, ICML '97. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 170–178.

Kosmopoulos, A., Partalas, I., Gaussier, E., Paliouras, G., Androutsopoulos, I., 2015. Evaluation measures for hierarchical classification: a unified view and novel approaches. Data Min. Knowl. Discov. 29, 820–865. https://doi.org/10.1007/s10618-014-0382-x

Kowsari, K., Brown, D., Heidarysafa, M., Jafari Meimandi, K., Gerber, M., Barnes, L., 2018. Web of Science Dataset. Mendeley Data V6. https://doi.org/10.17632/9rw3vkcfy4.6

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D., 2019. Text Classification Algorithms: A Survey. Information. 10. https://doi.org/10.3390/info10040150

Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D., 2017. Temporal Convolutional Networks for Action Segmentation and Detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1003–1012. https://doi.org/10.1109/CVPR.2017.113

Lewis, D.D., Yang, Y., Rose, T.G., Li, F., 2004. RCV1: A New Benchmark Collection for Text Categorization Research. J Mach Learn Res 5, 361–397.

Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P.S., He, L., 2022. A Survey on Text Classification: From Traditional to Deep Learning. ACM Trans. Intell. Syst. Technol. 13, 31:1-31:41. https://doi.org/10.1145/3495162

Liu, J., Chang, W.-C., Wu, Y., Yang, Y., 2017. Deep Learning for Extreme Multi-Label Text Classification, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17. Association for Computing Machinery, pp. 115–124. https://doi.org/10.1145/3077136.3080834

Lyubinets, V., Boiko, T., Nicholas, D., 2018. Automated Labeling of Bugs and Tickets Using Attention-Based Mechanisms in Recurrent Neural Networks, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). pp. 271–275. https://doi.org/10.1109/DSMP.2018.8478511

Madsen, A., Reddy, S., Chandar, S., 2022. Post-hoc Interpretability for Neural NLP: A Survey. ACM Comput. Surv. 55, 155:1-155:42. https://doi.org/10.1145/3546577

Mao, Y., Tian, J., Han, J., Ren, X., 2019. Hierarchical Text Classification with Reinforced Label Assignment, in: Inui, K., Jiang, J., Ng, V., Wan, X. (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November

3-7, 2019. Association for Computational Linguistics, pp. 445–455. https://doi.org/10.18653/v1/D19-1042

Marcuzzo, M., Zangari, A., Schiavinato, M., Giudice, L., Gasparetto, A., Albarelli, A., 2022. A multi-level approach for hierarchical Ticket Classification, in: Proceedings of the Eighth Workshop on Noisy User-Generated Text (W-NUT 2022). Presented at the WNUT 2022, Association for Computational Linguistics, Gyeongju, Republic of Korea, pp. 201–214.

Mencía, E.L., Fürnkranz, J., 2008. Efficient Pairwise Multilabel Classification for Large-Scale Problems in the Legal Domain, in: Proceedings of the 2008th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II, ECMLPKDD'08. Springer-Verlag, Berlin, Heidelberg, pp. 50–65. https://doi.org/10.1007/978-3-540-87481-2_4

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed Representations of Words and Phrases and their Compositionality, in: Advances in Neural Information Processing Systems 26, NIPS'13. Curran Associates, Inc., pp. 3111–3119.

Ni, J., Li, J., McAuley, J., 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 188–197. https://doi.org/10.18653/v1/D19-1018

Peng, H., Li, J., He, Y., Liu, Y., Bao, M., Wang, L., Song, Y., Yang, Q., 2018. Large-Scale Hierarchical Text Classification with Recursively Regularized Deep Graph-CNN, in: Proceedings of the 2018 World Wide Web Conference, WWW '18. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, pp. 1063–1072. https://doi.org/10.1145/3178876.3186005

Pennington, J., Socher, R., Manning, C., 2014. GloVe: Global Vectors for Word Representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162

Punera, K., Ghosh, J., 2008. Enhanced Hierarchical Classification via Isotonic Smoothing, in: Proceedings of the 17th International Conference on World Wide Web, WWW '08. Association for Computing Machinery, New York, NY, USA, pp. 151–160. https://doi.org/10.1145/1367497.1367518

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving Language Understanding by Generative Pre-Training.

Rios, A., Kavuluru, R., 2018. Few-Shot and Zero-Shot Multi-Label Learning for Structured Label Spaces, in: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Presented at the EMNLP 2018, Association for Computational Linguistics, Brussels, Belgium, pp. 3132–3142. https://doi.org/10.18653/v1/D18-1352

Salton, G., McGill, M.J., 1986. Introduction to Modern Information Retrieval. McGraw-Hill, Inc., USA.

Saranya, A., Subhashini, R., 2023. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decis. Anal. J. 7, 100230. https://doi.org/10.1016/j.dajour.2023.100230

Sharma, P., Shakya, A., Joshi, B., Panday, S.P., 2022. Hierarchical Multi Label Classification of News Articles Using RNN, CNN and HAN, in: Senjyu, T., Mahalle, P.N., Perumal, T., Joshi, A. (Eds.), ICT with Intelligent Applications, Smart Innovation, Systems and Technologies. Springer, Singapore, pp. 499–506. https://doi.org/10.1007/978-981-16-4177-0_50

Silla, C.N., Freitas, A.A., 2011. A survey of hierarchical classification across different application domains. Data Min. Knowl. Discov. 22, 31–72. https://doi.org/10.1007/s10618-010-0175-9

Stein, R.A., Jaques, P.A., Valiati, J.F., 2019. An analysis of hierarchical text classification using word embeddings. Inf Sci 471, 216–232. https://doi.org/10.1016/j.ins.2018.09.001

Sun, A., Lim, E.-P., 2001a. Hierarchical Text Classification and Evaluation, in: Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM '01. IEEE Computer Society, USA, pp. 521–528.

Sun, A., Lim, E.-P., 2001b. Hierarchical Text Classification and Evaluation, in: Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM '01. IEEE Computer Society, USA, pp. 521–528.

Sun, A., Lim, E.-P., Ng, W.-K., 2003. Hierarchical Text Classification Methods and Their Specification, in: Cooperative Internet Computing. Springer US, Boston, MA, pp. 236–256. https://doi.org/10.1007/978-1-4615-0435-1_14

Sun, A., Lim, E.-P., Ng, W.-K., Srivastava, J., 2004. Blocking reduction strategies in hierarchical text classification. IEEE Trans. Knowl. Data Eng. 16, 1305–1308. https://doi.org/10.1109/TKDE.2004.50

Vaswani, A., Aggarwal, G., Netrapalli, P., Hegde, N.G., 2022. All Mistakes Are Not Equal: Comprehensive Hierarchy Aware Multi-label Predictions (CHAMP). ArXiv Prepr. https://doi.org/10.48550/ARXIV.2206.08653

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Wang, Z., Wang, P., Huang, L., Sun, X., Wang, H., 2022. Incorporating Hierarchy into Text Encoder: a Contrastive Learning Approach for Hierarchical Text Classification, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Dublin, Ireland, pp. 7109–7119. https://doi.org/10.18653/v1/2022.acl-long.491

Yan, J., Mu, L., Wang, L., Ranjan, R., Zomaya, A.Y., 2020. Temporal Convolutional Networks for the Advance Prediction of ENSO. Sci. Rep. 10, 8055. https://doi.org/10.1038/s41598-020-65070-5

Yu, C., Shen, Y., Mao, Y., 2022. Constrained Sequence-to-Tree Generation for Hierarchical Text Classification, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22. Association for Computing Machinery, New York, NY, USA, pp. 1865–1869. https://doi.org/10.1145/3477495.3531765

Zhang, X., Xu, J., Soh, C., Chen, L., 2022. LA-HCN: Label-based Attention for Hierarchical Multi-label Text Classification Neural Network. Expert Syst Appl 187, 115922. https://doi.org/10.1016/j.eswa.2021.115922

Zhang, X., Zhao, J., LeCun, Y., 2015. Character-level Convolutional Networks for Text Classification, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Zhang, Y., Shen, Z., Dong, Y., Wang, K., Han, J., 2021. MATCH: Metadata-Aware

Text Classification in A Large Hierarchy, in: Proceedings of the Web Conference 2021, WWW '21. Association for Computing Machinery, New York, NY, USA, pp. 3246–3257. https://doi.org/10.1145/3442381.3449979

Zhang, Y., Wallace, B.C., 2017. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP). Asian Federation of Natural Language Processing, Taipei, Taiwan, pp. 253–263.

Zhao, R., Wei, X., Ding, C., Chen, Y., 2021. Hierarchical Multi-label Text Classification: Self-adaption Semantic Awareness Network Integrating Text Topic and Label Level Information, in: Qiu, H., Zhang, C., Fei, Z., Qiu, M., Kung, S.-Y. (Eds.), Knowledge Science, Engineering and Management. Springer International Publishing, Cham, pp. 406–418. https://doi.org/10.1007/978-3-030-82147-0_33

Zhou, J., Ma, C., Long, D., Xu, G., Ding, N., Zhang, H., Xie, P., Liu, G., 2020. Hierarchy-Aware Global Model for Hierarchical Text Classification, in: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics, pp. 1106–1117. https://doi.org/10.18653/v1/2020.acl-main.104