Hyperbolic Hierarchical Topic-based Keyphrase Generation

1st Wu Zhuang
Civil Aviation University of China
Tianjin, China
2023052082@cauc.edu.cn
2nd Heng Yu
Civil Aviation University of China
Tianjin, China
2022051038@cauc.edu.cn
3rd Yafu Li
Civil Aviation University of China
Tianjin, China
2022051037@cauc.edu.cn

***Abstract*—Keyphrases capture core information and enable precise retrieval. Documents are often organized hierarchically, and keyphrases concisely summarize high-level topics. Thus, accurate keyphrase prediction requires a deep understanding of such hierarchies for effective guidance. However, existing methods integrating hierarchical topic information into neural keyphrase generation remain confined to Euclidean space, limiting their ability to capture hierarchical structures. To address this, we propose Hyper-HTKG, a hyperbolic hierarchical topic-based keyphrase generation method that effectively leverages hierarchical topics to enhance performance. Specifically, we introduce a novel hyperbolic sequence generation approach with two main modules: a hyperbolic hierarchical topic model that learns latent topic trees across the document corpus, and a hyperbolic keyphrase generation model that produces keyphrases guided by hierarchical topics. Both modules are jointly trained to learn complementary information. To our knowledge, this is the first study on hyperbolic hierarchical topic-based keyphrase generation. Experiments on five benchmarks show Hyper-HTKG outperforms seven baselines.**

***Index Terms*—hyperbolic keyphrase generation model, hyperbolic hierarchical topic model, joint learning.**

I. Introduction

Keyphrase prediction is a fundamental NLP task that aims to automatically extract or generate concise keyphrases summarizing a document's core theme. Keyphrases offer high-level semantic abstraction and support downstream applications such as document indexing, retrieval [1], summarization [2], and sentiment analysis [3].

Despite ongoing efforts, keyphrase prediction remains challenging, with current models yielding suboptimal results. A major obstacle lies in determining whether candidate phrases genuinely capture the central ideas of a document—a task that demands deep semantic understanding beyond surface-level text matching.

Existing keyphrase prediction methods fall into two broad categories: conventional extraction techniques and deep generative approaches. Extraction methods are limited to identifying keyphrases that appear verbatim in the document, while deep generative models can produce both present and absent keyphrases.

Recent years have seen various topic-enhanced methods emerge across both paradigms. Extraction-based approaches include topic clustering [4] and topic-aware graph ranking [5]. Another line of research jointly learns latent topic representations in a flat structure [6].

Despite promising performance, most topic-aware models share a limiting premise: topics are treated as independent and organized in a flat hierarchy. To address this, a recent study [7] integrates neural hierarchical topic modeling into deep keyphrase generation, but its ability to capture hierarchical structures is constrained by Euclidean space. Although [8] introduces a hyperbolic topic-based method, it remains limited to flat topic representations.

Documents often contain multiple topics organized hierarchically rather than in a flat structure. As illustrated in Figure 1, the topic tree captures the hierarchical organization and semantic connections within a document. Building on this, we propose a strategy that first learns the document's topic hierarchy and then selects the most representative keyphrases from this structure as output, ensuring comprehensive coverage and high-level thematic abstraction. To our knowledge, no existing work leverages hierarchical topic structures in hyperbolic space to guide keyphrase generation.

Most current deep learning methods directly encode and decode documents without explicitly modeling topic structures. In contrast, our approach first models the hierarchical topic structure in hyperbolic space, then jointly trains the keyphrase generation model and hierarchical topic model to inform the generation process.

Our main contributions can be summarized as follows:

(1)To the best of our knowledge, this is the first attempt to leverage the hierarchical topics to guide deep keyphrase

generation in hyperbolic space.

(2) We propose a novel hyperbolic hierarchical topic-based keyphrase generation model that not only effectively captures the long and strong dependencies of a document, but also utilizes the hierarchical topics discovered to guide keyphrase generation.

(3)We compare our Hyper-HTKG method against six seq2seq keyphrase generation methods and a Sci-LLM(large language model used for scientific articles). Comprehensive experimental results demonstrate that our proposed method consistently outperforms the current SOTA (state-of-the-art) baseline methods across five publicly available datasets.

Example (#203 in KP20k)
Title: Fixed points of correspondences defined on cope metric spaces
Abstract: In the present note, we investigate the fixed points of correspondences defined on one metric spaces satisfying a conditionally contractive condition.
Gold: fixed point; cone metric space; correspondence; banach lattice*
Hyper-HTKG: **fixed point; cone metric space; correspondence**
CatSeqTG: fixed points; cone metric spaces; **correspondence**; fixed point matching;cone metric
ExHiRD-h: **fixed point**; **cone metric space**; point set; cone; <digit>
One2Set: **cone metric space**; **fixed point**; fixed point of correspondences; conditionally contractive mapping

Fig. 1: An example of predicted keyphrases by different approaches. Author-assigned (i.e., Gold) keyphrases and correctly-predicted keyphrases are shown in bold.

II. Related work

A. Deep Keyphrase Generation

The success of sequence-to-sequence (Seq2Seq) models in tasks like machine translation opened new avenues for keyphrase generation. CopyRNN [9] first employed an attentional Seq2Seq framework with a copying mechanism to generate both present and absent keyphrases, shifting the task from extraction to true generation.

We observe that most existing deep learning-based keyphrase prediction methods fail to incorporate latent hierarchical topic information into the Seq2Seq framework.

B. Topic-based Keyphrase Generation

Early research [10] combined traditional topic models like LDA with keyphrase extraction, using topic distribution as an external feature to re-rank candidate phrases from statistical or graph-based methods.

With the rise of neural networks, research has shifted toward end-to-end joint learning of topics and keyphrases by integrating neural topic modules with sequence generation models, conditioning generation on the document's global topic distribution. This enables models to generate absent keyphrases aligned with the document's thematic content, deriving abstract semantics from topic information.

C. Hyperbolic Representation

A growing number of studies show that many types of complex data exhibit non-Euclidean structures [11]. Recently, hyperbolic embedding methods have been proposed to learn latent representations of hierarchical data, yielding encouraging results. In NLP, hyperbolic representation learning has been successfully applied to word embeddings [12] and sentence representations [13]. Moreover, hyperbolic geometry has been integrated into advanced deep learning frameworks such as hyperbolic neural networks [14] and hyperbolic attention networks [15].

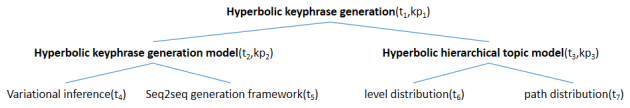

Fig. 2: The hierarchical topic tree structure of this paper.Three high-level topic(t1, t2, t3) description are selected as keyphrases (kp) finally.

III. Methodology

A. Problem Definition

Given a corpus of documents $D = \{d_i\}_{i=1}^{|D|}$, where each document d is a word sequence $X = (x_1, \ldots, x_{T_n})$ of length T_d, the goal of keyphrase generation is to produce a set of keyphrases $K = \{p_j\}_{j=1}^{|K|}$, with each keyphrase p as a word sequence $Y = (y_1, ..., y_{|P|})$. Following existing deep text generation models, X and Y denote the input document and its keyphrase sequence, respectively.

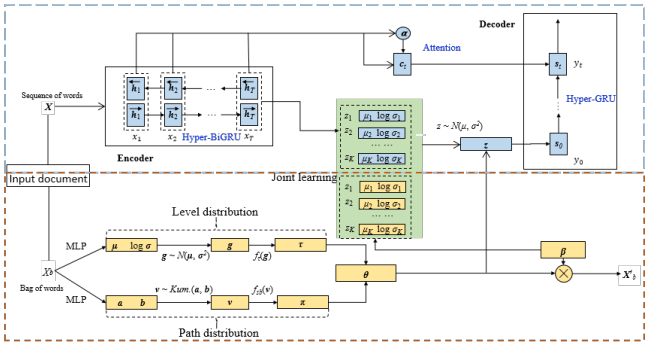

Fig. 3: The overall architecture of Hyper-HTKG .

B. Model Architecture

The overall architecture of our method is shown in Figure 3, consisting of two main modules. The Hyperbolic Hierarchical Topic Model learns a hierarchical topic distribution over a topic tree for each word occurrence in hyperbolic space, capturing the hierarchical semantic structure of the text. The Hyperbolic Keyphrase Generation Model then uses this distribution as guidance to generate keyphrases that are semantically aligned with the topic hierarchy and cover multi-granular information under hyperbolic geometric constraints. This framework naturally integrates hierarchical topic structures into keyphrase generation, enhancing semantic representational capacity and interpretability, and offers a new structure-aware,

geometrically interpretable direction for keyphrase generation tasks.

1) Hyperbolic Hierarchical Topic Model: We explicitly incorporate hierarchical topical information into keyphrase generation by adapting hierarchical topic modeling [16] to hyperbolic space to discover latent hierarchical topics.

Recent hierarchical topic models like TSNTM [17] and nTSNTM [18] employ autoencoding variational inference, offering better scalability and differentiable learning. These approaches typically use the nested Chinese Restaurant Process (nCRP) [19] to construct topic trees with unbounded branches and depths, where the distribution over tree paths is generated via a nested stick-breaking process:

$$v_k \sim \text{Beta}(1, \gamma) \tag{1}$$

$$\pi_k = \pi_{\text{par}(k)} v_k \prod_{j=1}^{k-1} (1 - v_j) \tag{2}$$

and the level distribution from a stick-breaking process:

$$\eta_l \sim \text{Beta}(1, \alpha) \tag{3}$$

$$\theta_l = \eta_l \prod_{j=1}^{l-1} (1 - \eta_j) \tag{4}$$

where $k \in \{1, \ldots, K\}$ and par(k) denote $k - th$ and its parent. $l \in \{1, \ldots, L\}$, v_k and η_l are stick proportions.

Given the topic distribution θ of document d and the topic-word distribution β, the hierarchical topic-guided Gaussian mixture model can be expressed as follows:

$$p(z|X_b) = \sum_{k=1}^{K} \theta_k(X_b) N(z_k; \mu_k(X_b), \sigma_k^2(X_b)) \tag{5}$$

Building on the approach introduced in prior work [20], the topic-word distribution—denoting the probability distribution over vocabulary terms for a given topic k—is computed explicitly as follows:

$$\beta_k = softmax(E_w \cdot e_k^T) \tag{6}$$

where e_k and E_w are the embedding of topic k and all the words.

Within the Auto-Encoding Variational Bayes (AEVB) framework, the evidence lower bound (ELBO) for the log-likelihood of document d is derived as:

$$\begin{aligned} \mathcal{L}_{ht} =& E_{q((\pi|X_b))}[\sum \log(\theta \cdot \beta)] \\ & - KL[q(\pi|X_b)||p(\pi)] - KL[q(\tau|X_b)||p(\tau)] \end{aligned} \tag{7}$$

where $q(\pi|X_b)$ and $q(\tau|X_b)$ are posteriors modeled by the inference network, and $p(\pi)$ and $p(\tau)$ are the priors.

2) Hyperbolic Keyphrase Generation Guided by Hierarchical Topic: Unlike conventional Seq2Seq keyphrase generation models such as SEG-Net [21] and CopyRNN [9], our model operates within a hyperbolic Seq2Seq architecture. Specifically, we introduce a latent variable informed by the hierarchical topic model that captures the document's underlying topic structure and serves as a global semantic guide during generation, enabling the model to effectively leverage high-level thematic information.

We use a bidirectional GRU in hyperbolic space (Hyper-BiGRU) [22] as the encoder:

$$\mathbf{h}_1, \ldots, \mathbf{h}_T = \text{Hyper-BiGRU}(x_1, \ldots, x_T) \tag{8}$$

Each contextualized vector $\mathbf{h}_t$ encodes context information for the t-th word, and the last hidden state $\mathbf{h}_T$ is used to compute the latent topic variable $\mathbf{z}$.

Topic information is then incorporated into latent variables, with each topic drawn from a topic-specific multivariate Gaussian distribution:

$$p(\mathbf{z}|\mathbf{X}) = \sum_{k=1}^{K} \theta_k(\mathbf{X}_b) \mathcal{N}\left(\mathbf{z}; \widehat{\boldsymbol{\mu}}_k(\mathbf{X}), \widehat{\boldsymbol{\sigma}}_k^2(\mathbf{X})\right) \tag{9}$$

where θ_k is the usage of topic k in the document, computed by our hyperbolic-HTM. We use fully connected layers to estimate $\widehat{\boldsymbol{\mu}}_k$ and $\log \widehat{\boldsymbol{\sigma}}_k$:

$$\widehat{\boldsymbol{\mu}}_k = \mathbf{W}_{\mu_k} \mathbf{h}_T + \mathbf{b}_{\mu_k} \tag{10}$$

$$\log \widehat{\boldsymbol{\sigma}}_k = \mathbf{W}_{\sigma_k} \mathbf{h}_T + \mathbf{b}_{\sigma_k} \tag{11}$$

The latent topic variable $\mathbf{z}$ is obtained using the reparameterization trick.

Finally, given a source document $\mathbf{X}$ and latent topic variable $\mathbf{z}$, the keyphrase $\mathbf{Y}$ is generated by:

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{t=1}^{|\mathbf{Y}|} p(y_t|\mathbf{Y}_{<t}, \mathbf{z}, \mathbf{X})\, p(\mathbf{z}|\mathbf{X}) \tag{12}$$

where $\mathbf{Y}_{<t} = (y_1, \ldots, y_{t-1})$. The decoder is a forward GRU that predicts the next word y_t based on its hidden state s_t, conditioned on y_{t-1} and the context vector $\mathbf{c}_t$:

$$s_t = \text{GRU}_f(y_{t-1}, s_{t-1}, \mathbf{c}_t) \tag{13}$$

$$p(y_t|\mathbf{Y}_{<t}, \mathbf{z}, \mathbf{X}) = g(y_{t-1}, s_t, \mathbf{c}_t) \tag{14}$$

Here, $\mathbf{c}_t = \sum_i \alpha_{ti} \mathbf{h}_i$ is the vector of attention source context, and $g(\cdot)$ is a non-linear function that outputs the probability of y_t'. $\mathbf{z}$ initializes the initial state of the decoder s_0.

We minimize the cross-entropy loss:

$$\mathcal{L}_{kg} = -\sum_{t=1}^{N} \log\left(p(y_t|\mathbf{X}, \mathbf{Y}_{<t}, \mathbf{z})\right) \tag{15}$$

where N is the length of target keyphrase.

3) Joint Learning: Given that keyphrase generation and topic modeling both aim to identify and condense core semantic information from documents, we propose a co-training framework that jointly optimizes both tasks. This enables bidirectional knowledge transfer and semantic complementarity between the two modules during training, enhancing the model's ability to capture multi-level semantic structures.

To align the hierarchical topic-guided Gaussian mixture with the corresponding distribution derived from the hyperbolic topic model, we introduce a third loss term, an inconsistency loss $\mathcal{L}_{ic}$ as the KL divergence between $p(\mathbf{z}|\mathbf{X}_b) =$

$\sum_{i=1}^{K} \theta_i \mathcal{N}(\mu_i, \sigma_i^2)$ and $p(\mathbf{z}|\mathbf{X}) = \sum_{i=1}^{K} \theta_i \mathcal{N}(\widehat{\mu}_i, \widehat{\sigma}_i^2)$, their Kullback–Leibler (KL) divergence is upper-bounded by:

$$\begin{aligned} \mathcal{L}_{ic} &= \mathrm{KL}(p(\mathbf{z}|\mathbf{X}_b) \| p(\mathbf{z}|\mathbf{X})) \\ &\leq \sum_{i=1}^{K} \theta_i \, \mathrm{KL}(\mathcal{N}(\mu_i, \sigma_i^2) \| \mathcal{N}(\widehat{\mu}_i, \widehat{\sigma}_i^2)) \end{aligned} \tag{16}$$

where $\mathrm{KL}(\boldsymbol{\theta} \| \boldsymbol{\theta}) = 0$

The overall training objective is formulated as a linear combination of the three loss components, expressed as follows:

$$\mathcal{L} = \mathcal{L}_{ht} + \mathcal{L}_{kg} + \mathcal{L}_{ic} \tag{17}$$

IV. Experiment

A. Datasets

We use the KP20k dataset compiled by Meng et al. [9], a large-scale collection of scientific metadata from the computer science domain sourced from ACM, Springer, and IEEE. Widely recognized for its domain representativeness and annotation consistency, each instance includes the title and abstract of a scholarly article as source text, with author-assigned keywords as target keyphrases. Due to its scale and reliable annotations, KP20k has become a standard benchmark for keyphrase generation.

To evaluate generalizability and domain adaptability, we conduct additional experiments on four widely-used public datasets from the scientific domain—Inspec [23], Krapivin [24], NUS [25], and SemEval [26]—using the model trained on KP20k. This cross-dataset assessment examines robustness and transferability across variations in data distribution, textual style, and annotation practice. Table 1 presents key statistics for these five datasets, including number of documents (#Doc), number and percentage of present keyphrases (#PKp and %PKp), number and percentage of absent keyphrases (#AKp and %AKp), and average keyphrases per document (#Avg.Kp).

TABLE I: Summary of the datasets.

Dataset	#Doc	#PKp	%PKp	#AKp	%AKp	#Avg.Kp
KP20k	569K	1.6M	63.0	995K	37.0	5.28
Inspec	500	3602	73.6	1293	26.4	9.79
Krapivin	400	1297	55.6	1037	44.4	5.84
NUS	211	1191	52.2	1088	47.8	10.8
SemEval	100	612	42.4	831	57.6	14.43

B. Evaluation Metrics

To enable fair comparison with existing studies, we adopt $F1@k$ and $F1@O$ as our primary evaluation metrics.

$F1@k$ is computed from the top-k predicted keyphrases ($k = 5$ or 10), assuming a fixed output length. $F1@O$ [27] matches the number of predicted phrases to the ground-truth count per document, avoiding bias from fixed-length output.

Both metrics are harmonic means of precision and recall, offering a balanced measure of accuracy and coverage.

C. Experiment Setup

Following prior work [9] [27], we preprocess data with lowercase and tokenization. The encoder/decoder vocabulary consists of the 50,000 most frequent training words, while a separate bag-of-words vocabulary of the top 10,000 terms is used for the topic model.

In the hyperbolic hierarchical topic model, the hidden dimension is 256, with hyperparameters $\alpha_0 = 1$ and $\beta_0 = 10$. For keyphrase generation, word embeddings are 150-dimensional, the Hyper-BiGRU encoder has a hidden size of 150, and the forward GRU decoder uses 300 dimensions. Other hyperparameters follow established configurations.

We adopt the One2One generation paradigm [9] and optimize with Adam (learning rate 0.001, gradient clipping 1.0). Batch sizes are 1024 for the topic model and 128 for the generation model. Learning rate decays on validation plateau, with early stopping after three epochs.

To balance convergence, we pre-train the topic model for 100 epochs before joint training. During joint training, the KL weight is gradually increased from 0 to 1 to prevent posterior collapse. Inference uses beam search (width 50, max depth 6).

D. Performance Comparison

Tables 2 and 3 show the results for present and absent keyphrase prediction, respectively. For present keyphrases, Hyper-HTKG achieves the best or second-best performance on all five datasets, with gains of up to 1.6 F1@5 points on Inspec. For absent keyphrases, our model consistently outperforms all baselines across every dataset, with particularly large gains on NUS and SemEval. These results demonstrate that Hyper-HTKG effectively captures document semantics for both present and absent keyphrase generation.

E. Ablation Study

To evaluate the contribution of each component within our proposed Hyper-HTKG framework, we conduct an ablation study by comparing the full model against the following ablated variants: (1) w/o HS(hyperbolic space) where the hyperbolic space is replaced bu Euclidean space, (2) w/o HT(hierarchical topic) where the hierarchical topic model is replaced by the flat topic model [20], (3) w/o TG(topic guidance), where we directly use the hyperbolic keyphrase generation model to generate keyphrases without topic guidance obtained from the hyperbolic hierarchical topic model.

From the results shown in Table 4, we have the following observations: (1)Replacing the hyperbolic space with the conventional Euclidean space leads to a noticeable degradation in model performance. This comparison further underscores the essential role of hyperbolic space in effectively modeling hierarchical semantics. (2) Removal of topic guidance results in a significant performance drop on all datasets, indicating that topic guidance is effective information to improve keyphrase generation. (3) Replacing hierarchical topics with flat topics leads to performance drops across all datasets, indicating that hierarchical topic information is more effective than flat topics in improving keyphrase generation.

TABLE II: Present keyphrase prediction results on five datasets. The best performing score is highlighted in bold and the second best score is highlighted with underline.

Model	KP20k		Inspec		Krapivin		NUS		SemEval	
	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O
CopyRNN	31.7	33.5	24.4	29.0	30.5	32.5	37.6	40.6	31.8	**31.7**
CopyCNN	35.1	–	28.5	–	31.4	–	34.2	–	29.5	–
Sci-LLM	36.7	**38.8**	25.7	31.4	27.2	31.7	28.9	38.4	20.2	30.3
CatSeq	31.4	31.9	29.0	30.7	30.7	32.4	35.9	38.3	30.2	31.0
CatSeqTG	32.6	35.7	26.6	22.4	31.2	**34.7**	36.5	39.6	27.7	25.5
Exhird-h	31.1	37.4	25.3	28.9	28.4	30.6	–	–	29.2	26.6
One2Set	35.5	36.9	28.2	25.4	31.5	34.3	39.7	39.5	**34.0**	30.2
Ours	**37.4**	**38.8**	**30.6**	**31.7**	**31.7**	32.3	**40.9**	**41.7**	32.0	30.5

TABLE III: Absent keyphrase prediction results on five datasets. The best performing score is highlighted in bold and the second best score is highlighted with underline.

Model	KP20k		Inspec		Krapivin		NUS		SemEval	
	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O
CopyRNN	3.23	3.97	1.44	1.23	5.14	5.54	3.35	3.25	2.05	2.20
CatSeq	1.50	2.51	0.40	0.35	1.80	1.67	1.60	1.22	1.60	1.44
CatSeqTG	2.78	2.16	1.12	0.82	2.97	1.88	2.46	2.10	2.04	1.78
Exhird-h	1.57	3.22	1.03	1.52	2.06	2.49	–	–	1.51	1.14
One2Set	3.70	5.02	2.10	1.63	**5.53**	3.43	4.02	3.61	2.42	2.02
Ours	**4.20**	**5.27**	**2.24**	1.59	5.41	**5.82**	**4.96**	**4.82**	**2.99**	**2.67**

TABLE IV: Performance Comparison

	Model	KP20k		Inspec		Krapivin		NUS		SemEval	
		F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O	F1@5	F1@O
present	Ours	37.4	38.8	30.6	31.7	31.7	32.3	40.9	41.7	32.0	30.5
	w/o HS	35.5	36.8	29.3	29.9	28.3	28.8	37.6	39.1	31.0	29.3
	w/o HT	34.1	36.3	27.4	28.0	26.9	28.4	34.6	36.2	28.2	27.5
	w/o TG	33.0	34.3	25.5	27.6	26.1	27.8	32.4	33.6	27.1	26.4
absent	Ours	4.20	5.27	2.24	1.59	5.41	5.82	4.96	4.82	2.99	2.67
	w/o HS	3.12	3.64	1.99	1.54	3.68	4.01	3.52	3.93	2.31	2.32
	w/o HT	2.73	3.32	1.86	1.44	2.89	3.10	3.27	3.62	1.98	2.02
	w/o TG	1.99	2.20	1.46	1.27	2.05	2.26	2.07	2.42	1.69	1.80

V. Conclusion

In this study, we introduced Hyper-HTKG, a keyphrase generation model that integrates hierarchical topic modeling within hyperbolic space. By jointly learning latent topics and generating keyphrases, our approach captures the inherent semantic structure of documents. Experiments on five benchmark datasets show that Hyper-HTKG achieves state-of-the-art performance. In future work, we will perform statistical tests to more comprehensively evaluate the robustness of the proposed method. We will make our data/code available at www.Github.com upon acceptance.

References

[1] Xiaojun Wan and Jianguo Xiao. “Single document keyphrase extraction using neighborhood knowledge.” In: *AAAI*. Vol. 8. 2008, pp. 855–860.

[2] Lu Wang and Claire Cardie. “Domain-independent abstract generation for focused meeting summarization”. In: *the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2013, pp. 1395–1405.

[3] Gabor Berend. “Opinion Expression Mining by Exploiting Keyphrase Extraction”. In: *Proceedings of IJCNLP*. 2011.

[4] Maria Grineva, Maxim Grinev, and Dmitry Lizorkin. “Extracting key terms from noisy and multitheme doc-

uments". In: *Proceedings of the 18th International Conference on World Wide Web*. WWW '09. Madrid, Spain: Association for Computing Machinery, 2009, pp. 661–670. ISBN: 9781605584874. DOI: 10.1145/1526709.1526798. URL: https://doi.org/10.1145/1526709.1526798.

[5] Florian Boudin. "Unsupervised Keyphrase Extraction with Multipartite Graphs". In: *North American Chapter of the Association for Computational Linguistics*. 2018. URL: https://api.semanticscholar.org/CorpusID:4792796.

[6] Yue Wang et al. "Topic-Aware Neural Keyphrase Generation for Social Media Language". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by Anna Korhonen, David Traum, and Lluís Màrquez. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 2516–2526. DOI: 10.18653/v1/P19-1240. URL: https://aclanthology.org/P19-1240/.

[7] Yuxiang Zhang et al. "HTKG: Deep Keyphrase Generation with Neural Hierarchical Topic Guidance". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '22. Madrid, Spain: Association for Computing Machinery, 2022, pp. 1044–1054. ISBN: 9781450387323. DOI: 10.1145/3477495.3531990. URL: https://doi.org/10.1145/3477495.3531990.

[8] Yuxiang Zhang et al. "Hyperbolic Deep Keyphrase Generation". In: *ECML/PKDD*. 2022. URL: https://api.semanticscholar.org/CorpusID:257633299.

[9] Rui Meng et al. "Deep Keyphrase Generation". In: *Proceedings of ACL*. 2017, pp. 582–592.

[10] Adrien Bougouin, Florian Boudin, and Béatrice Daille. "Topicrank: Graph-based topic ranking for keyphrase extraction". In: *Proceedings of IJCNLP*. 2013.

[11] Michael M. Bronstein et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418.

[12] Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. "Poincar\'e glove: Hyperbolic word embeddings". In: *arXiv preprint arXiv:1810.06546* (2018).

[13] Bhuwan Dhingra et al. "Embedding text in hyperbolic spaces". In: *Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12)*. 2018, pp. 59–69.

[14] Maximillian Nickel and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations". In: *Advances in neural information processing systems* 30 (2017).

[15] Yiding Zhang et al. "Hyperbolic Graph Attention Network". In: *IEEE Transactions on Big Data* 8.6 (2022), pp. 1690–1701. DOI: 10.1109/TBDATA.2021.3081431.

[16] Ziye Chen et al. "Tree-structured topic modeling with nonparametric neural variational inference". In: *Proceedings of ACL*. 2021.

[17] Masaru Isonuma et al. "Tree-Structured Neural Topic Model". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Ed. by Dan Jurafsky et al. Online: Association for Computational Linguistics, July 2020, pp. 800–806. DOI: 10.18653/v1/2020.acl-main.73. URL: https://aclanthology.org/2020.acl-main.73/.

[18] Ziye Chen et al. "Tree-Structured Topic Modeling with Nonparametric Neural Variational Inference". In: *Proceedings 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)*. Ed. by Chengqing Zong et al. Online: Association for Computational Linguistics, Aug. 2021, pp. 2343–2353. DOI: 10.18653/v1/2021.acl-long.182. URL: https://aclanthology.org/2021.acl-long.182/.

[19] Joon Hee Kim et al. "Modeling topic hierarchies with the recursive chinese restaurant process". In: *Proceedings of CIKM*. 2012.

[20] Yishu Miao, Edward Grefenstette, and Phil Blunsom. "Discovering discrete latent topics with neural variational inference". In: *International conference on machine learning*. PMLR. 2017, pp. 2410–2419.

[21] Wasi Uddin Ahmad et al. "Select, Extract and Generate: Neural Keyphrase Generation with Layer-wise Coverage Attention". In: *Proceedings of ACL*. 2021.

[22] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *Proceedings of ICLR*. 2015.

[23] Anette Hulth and Beáta B Megyesi. "A study on automatically extracted keywords in text categorization". In: *Proceedings of ACL*. 2006.

[24] Mikalai Krapivin, Aliaksandr Autaeu, and Maurizio Marchese. *Large dataset for keyphrases extraction*. Tech. rep. University of Trento, 2009.

[25] Thuy Dung Nguyen and Min-Yen Kan. "Keyphrase extraction in scientific publications". In: *Proceedings of ICADL*. 2007.

[26] Su Nam Kim et al. "SemEval-2010 Task 5 : Automatic Keyphrase Extraction from Scientific Articles". In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. Ed. by Katrin Erk and Carlo Strapparava. Uppsala, Sweden: Association for Computational Linguistics, July 2010, pp. 21–26. URL: https://aclanthology.org/S10-1004/.

[27] Xingdi Yuan et al. "One Size Does Not Fit All: Generating and Evaluating Variable Number of Keyphrases". In: *Proceedings of ACL*. 2020.