A Preliminary Exploration of Extractive Multi-Document Summarization in Hyperbolic Space

Mingyang Song
mingyang.song@bjtu.edu.cn
Beijing Jiaotong University
Beijing, China

Yi Feng
21112027@bjtu.edu.cn
Beijing Jiaotong University
Beijing, China

Liping Jing*
lpjing@bjtu.edu.cn
Beijing Jiaotong University
Beijing, China

ABSTRACT

Summary matching is a recently proposed paradigm for extractive summarization. It aims to calculate similarities between candidate summaries and their corresponding document and extract summaries by ranking similarities. Due to natural languages often exhibiting the inherent hierarchical structures ingrained with complex syntax and semantics, the latent hierarchical structures between candidate summaries and their corresponding document should be considered when calculating the summary-document similarities. However, the above structural property is hard to model in the Euclidean space. Inspired by the above issues, we explore extractive summarization in the hyperbolic space and propose a new Hyperbolic Siamese Network for the matching-based extractive summarization (HyperSiameseNet). Specifically, HyperSiameseNet projects candidate summaries and their corresponding document representations from the Euclidean space to the Hyperbolic space and then models the summary-document similarities via the squared poincaré distance. Finally, the summary-document similarities are optimized by the margin-based triplet loss for extracting the final summary. The results on the Multi-News dataset have shown the superiority of our model HyperSiameseNet by comparing with the state-of-the-art baselines[1].

CCS CONCEPTS

• **Computing methodologies** → *Information extraction*; • **Information systems** → **Summarization**.

KEYWORDS

extractive summarization, hyperbolic space, siamese network

ACM Reference Format:
Mingyang Song, Yi Feng, and Liping Jing. 2022. A Preliminary Exploration of Extractive Multi-Document Summarization in Hyperbolic Space. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557538

*Corresponding author.
[1]https://github.com/MySong7NLPer/HyperSiameseNet

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557538

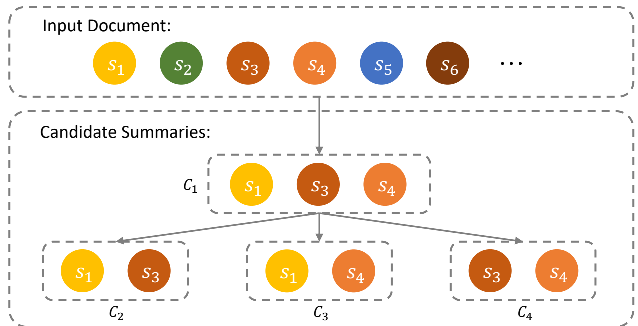

Figure 1: The latent hierarchical structures between candidates and the document. S_* indicates sentences in the source document and C_* denotes candidate summaries. A candidate summary with more sentences typically has more complex structures than one with fewer sentences.

1 INTRODUCTION

Document summarization has attracted much attention and has a wide-ranging application with the explosive growth of text information, which aims to condense a long document into a short version with salient information of the given document. Automatic document summarization models can be broadly divided into extractive and abstractive methods. Extractive summarization generates a summary by selecting salient sentences or phrases from the source document, while abstractive methods paraphrase and restructure sentences to compose a summary. We focus on extractive summarization in this work as it is faster and thus can generate more reliable summaries. Specifically, extractive summarization includes two types: sentence- and summary-level approaches. The former considers the semantics of each sentence independently, resulting in high redundancy. On the contrary, the latter considers the semantics of multiple sentences simultaneously, suffering from high computational complexity.

To take advantage of the sentence- and summary-level extractive summarization, Zhong et al. [29] propose MatchSum, a new matching-based paradigm, is the first to formulate extractive summarization as a semantic matching problem. MatchSum first encodes the document and candidate summaries by a Siamese-BERT architecture and then calculates and ranks the similarities to extract a summary. The above method represents text information and models the similarities between candidate summaries and their corresponding document in the Euclidean space. Although the above model in the Euclidean space has proved successful for the extractive summarization task, it still suffers from an inherent limitation:

the ability to consider the latent hierarchical structures when representing text information and calculating the similarities is bounded by the nature of flat geometry provided by the Euclidean space, as mentioned in recent research [20].

Recently, constructing neural networks in the hyperbolic space have been developed to capture the latent hierarchical nature of data and demonstrated encouraging results [20]. To efficiently utilize the hyperbolic properties in downstream tasks, researchers propose some advanced hyperbolic deep networks [10]. Many recent studies show the superiority of the hyperbolic space for many natural language processing tasks [12, 24].

Motivated by the above observations, we explore the matching-based extractive summarization in the hyperbolic space and propose a hyperbolic siamese network for the matching-based extractive summarization. Specifically, we first obtain candidate summaries and their corresponding document representations by BERT and then map them to the same hyperbolic space. Next, we model the latent hierarchical structures by calculating the relevance between candidate summaries and the document via the squared poincaré distance as the similarities. Finally, our model is minimized by the margin-based triplet loss to extract the best candidate as the summary. This is the first attempt to explore extractive summarization in the hyperbolic space to the best of our knowledge. The experimental results on the Multi-News dataset have shown the superiority of the proposed model by comparing with the recent baselines.

2 PRELIMINARIES

Hyperbolic space is an important concept in hyperbolic geometry, which is considered as a special case in the Riemannian geometry [14]. Before introducing our model, this section briefly presents the basic information of hyperbolic space.

In a traditional sense, hyperbolic spaces are not vector spaces; one cannot use standard operations such as summation, multiplication, etc. To remedy this problem, one can utilize the formalism of Möbius gyrovector spaces allowing to generalize many standard operations to hyperbolic spaces [15]. The hyperbolic space can be described via *Riemannian geometry* [14]. Similarly to the previous studies [10, 20, 25], we adopt the *Poincaré ball* and use an additional hyper-parameter c which modifies the curvature of *Poincaré ball*; it is then defined as $\mathbb{D}_c^n = \{\mathbf{x} \in \mathbb{R}^n : c\|\mathbf{x}\|^2 < 1, c \geq 0\}$. The corresponding conformal factor now takes the form $\lambda_{\mathbf{x}}^c := \frac{2}{1-c\|\mathbf{x}\|^2}$. In practice, the choice of c allows one to balance hyperbolic and Euclidean geometries, which is made precise by noting that with $c \to 0$, all the formulas discussed below take their usual Euclidean form. In this paper, we use the standard Poincaré ball $\mathbb{D}_c^n$ with $c = 1$. We restate the definitions of fundamental mathematical operations for the generalized Poincaré ball model and refer readers to [10] for more details. Next, we give details of the closed-form formulas of several Möbius operations.

Möbius Addition. For a pair $\mathbf{x}, \mathbf{y} \in \mathbb{D}_c^n$, the *Möbius addition* is defined as,

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2}. \tag{1}$$

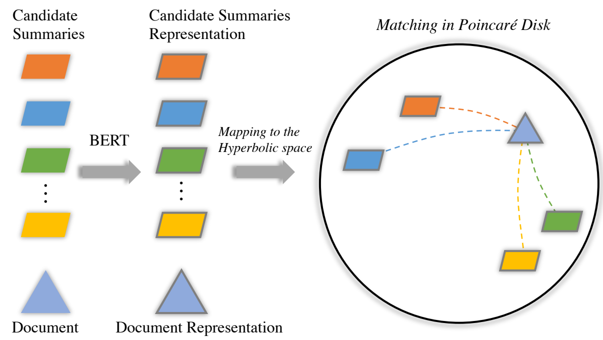

Figure 2: The model architecture of HyperSiameseNet.

Möbius Matrix-vector Multiplication. For a linear map $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^m$ and $\forall \mathbf{x} \in \mathbb{D}_c^n$, if $\mathbf{M}\mathbf{x} \neq 0$, then the *Möbius matrix-vector multiplication* is defined as,

$$\mathbf{M} \otimes_c \mathbf{x} = (\frac{1}{\sqrt{c}})\tanh(\frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|}\tanh^{-1}(\|\sqrt{c}\mathbf{x}\|))\frac{\mathbf{M}\mathbf{x}}{\|\mathbf{M}\mathbf{x}\|}, \tag{2}$$

where $\mathbf{M} \otimes_c \mathbf{x} = 0$ if $\mathbf{M}\mathbf{x} = 0$.

Poincaré Distance. The induced distance function for a pair $\mathbf{x}, \mathbf{y} \in \mathbb{D}_c^n$ is defined as,

$$d_c(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|). \tag{3}$$

Note that with $c = 1$ one recovers the geodesic distance, while with $c \to 0$ we obtain the Euclidean distance $\lim_{c\to 0} d_c(\mathbf{x}, \mathbf{y}) = 2\|\mathbf{x} - \mathbf{y}\|$.

Exponential and Logarithmic Maps. To perform operations in the hyperbolic space, one first needs to define a mapping function from $\mathbb{R}^n$ to $\mathbb{D}_c^n$ to map Euclidean vectors to the hyperbolic space. Let $T_{\mathbf{x}}\mathbb{D}_c^n$ denote the *tangent space* of $\mathbb{D}_c^n$ at $\mathbf{x}$. The *exponential map* $\exp_{\mathbf{x}}^c(\cdot) : T_{\mathbf{x}}\mathbb{D}_c^n \to \mathbb{D}_c^n$ for $\mathbf{v} \neq 0$ is defined as:

$$\exp_{\mathbf{x}}^c(\mathbf{v}) = \mathbf{x} \oplus_c (\tanh(\sqrt{c}\frac{\lambda_{\mathbf{x}}^c\|\mathbf{v}\|}{2})\frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|}). \tag{4}$$

As the inverse of $\exp_{\mathbf{x}}^c(\cdot)$, the *logarithmic map* $\log_{\mathbf{x}}^c(\cdot) : \mathbb{D}_c^n \to T_x\mathbb{D}_c^n$ for $\mathbf{y} \neq \mathbf{x}$ is defined as:

$$\log_{\mathbf{x}}^c(\mathbf{y}) = \frac{2}{\sqrt{c}\lambda_{\mathbf{x}}^c}\tanh^{-1}(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|)\frac{-\mathbf{x} \oplus_c \mathbf{y}}{\| - \mathbf{x} \oplus_c \mathbf{y}\|} \tag{5}$$

3 HYPERBOLIC SIAMESE NETWORK

Given the input document $\mathcal{D} = \{w_1, ..., w_i, ..., w_N\}$, HyperSiameseNet aims to calculate and rank the importance score of candidate summaries $\mathbf{C} = \{C_1, C_2, ..., C_{|\mathbf{C}|}\}$ depends on their corresponding document. Inspired by the previous work [29], to avoid the curse of combination, we prune the given document in sentence-level by BERTEXT [18] and construct candidate summaries.

3.1 Hyperbolic Encoder

The tree-likeness of the hyperbolic space [13] makes it natural to embed hierarchical structures in natural language [6]. Previous work has been shown that any finite tree can be embedded with arbitrary low distortion into the Poincaré ball while the distances are approximately preserved [22]. Conversely, it is difficult to perform such embedding in the Euclidean space even with unbounded dimensionality [21]. Recent work [3] verify that mapping the pre-trained word embeddings to the hyperbolic space can better model the hierarchical structures hidden in natural language.

Inspired by the above phenomena, we first adopt the Siamese-BERT [7] to obtain the document embedding $\mathcal{H}_d$, the ground-truth summary embedding $\mathcal{H}_s$, the k-th candidate summary embedding $\mathcal{H}_{c_k}$ and then map the obtained embeddings onto the hyperbolic space as follows,

$$\tilde{\mathcal{H}_d} = \exp_{\mathbf{0}}^c(\mathbf{W}_d\mathcal{H}_d), \tag{6}$$

$$\tilde{\mathcal{H}_s} = \exp_{\mathbf{0}}^c(\mathbf{W}_s\mathcal{H}_s), \tag{7}$$

$$\tilde{\mathcal{H}}_{c_k} = \exp_{\mathbf{0}}^c(\mathbf{W}_c\mathcal{H}_{c_k}), \tag{8}$$

where $\tilde{\mathcal{H}_d}, \tilde{\mathcal{H}_s}, \tilde{\mathcal{H}}_{c_k}$ denotes the hyperbolic document, ground-truth summary, candidate summary representation respectively. Here, $\mathbf{W}_d \in \mathbb{R}^{e*r}$, $\mathbf{W}_s \in \mathbb{R}^{e*r}$, and $\mathbf{W}_c \in \mathbb{R}^{e*r}$ indicate the linear transformation in euclidean space, e being the dimension of contextualized embeddings and r being the rank of hyperbolic representations, that projects the distributed representations to the tangent space.

3.2 Interaction in the Hyperbolic Space

The primary objective of summary matching is to build relevance from the candidate summaries to their corresponding document. To capture the structure-aware similarities, we model the relevance between candidate summaries and their corresponding document as follows, First, we calculate the textual similarity between candidate summary $\tilde{\mathcal{H}}_{c_k}$ and its corresponding document $\tilde{\mathcal{H}_d}$,

$$f_c(\mathcal{D}, C_i) = -d_c^2(\tilde{\mathcal{H}_d}, \tilde{\mathcal{H}}_{c_k}), \tag{9}$$

where $f_c(\mathcal{D}, C_i)$ can be considered the summary-document similarity score by matching the i-th candidate summary with its corresponding document. Here, $d_c^2(\cdot)$ indicates the squared poincaré distance. Specifically, since the candidate summary and the document are embedded by the same pre-trained language model, similar content will be placed closer in the hyperbolic space and their product will produce smaller scores.

3.3 Training and Inference

To learn representations and match summary, we adopt a margin-based triplet loss to optimise the model parameters:

$$\mathcal{L}_1 = \max(0, f_c(\mathcal{D}, \mathbf{C}) - f_c(\mathcal{D}, C^*) + \gamma_1) \tag{10}$$

where $\mathbf{C}$ indicates the candidate summary in $\mathcal{D}$ and γ_1 is a margin value. Through the above optimization objective, the ground-truth summary C^* should be semantically closest to the given document $\mathcal{D}$. Furthermore, another criterion for designing our loss function is that the candidate pair with a higher ranking disparity should have a wider margin. Therefore, we adopt a pairwise margin loss for ranking each candidate summary,

$$\mathcal{L}_2 = \max(0, f_c(\mathcal{D}, C_j) - f_c(\mathcal{D}, C_i) + (j - i) * \gamma_2) \quad (i < j), \tag{11}$$

where C_i denotes the candidate summary ranked i and γ_2 is a hyper-parameter used to distinguish between good and bad candidate summaries. Specifically, we arrange all candidate summaries in descending order of ROUGE scores with the ground-truth summary in the training phase. Finally, we combine the above two loss functions, $\mathcal{L} = \delta\mathcal{L}_1 + (1 - \delta)\mathcal{L}_2$, where δ indicates the balance factor of the above two loss. The primary idea is to provide the gold summary with the highest matching score while also giving a better candidate summary a greater score when compared to the unqualified candidate summary.

In the inference phase, we regard extractive summarization as a text matching task to find the best candidate summary among all candidates C_* extracted from the given document $\mathcal{D}$.

4 EXPERIMENTS AND RESULTS

4.1 Dataset and Evaluation Metric

Multi-News is a multi-document news summarization dataset with a relatively long summary. Following the previous work [29], we use the truncated version and concatenate the source documents as a single input in all experiments.

In this paper, we use ROUGE [17] to evaluate the predicted summary in our experiments, including ROUGE-1 (R-1), ROUGE-2 (R-2), and ROUGE-L (R-L). Furthermore, to observe the whole improvement of performance, we calculate the mean value (R-AVG.) of R-1, R-2, and R-L evaluation metrics.

4.2 Implementation Detail

We use the base version of BERT to implement our models in all experiments. RiemannianAdam[1] optimizer with warming-up is used and the learning rate schedule follows [26] as, $lr = 2e^{-3} \cdot \min(\text{step}^{-0.5}, \text{step} \cdot \text{wm}^{-1.5})$, where each step is a batch size of 16 and wm indicates warm-up steps of 10,000. To prevent the performance degradation [29], we choose $\gamma_1 = 0$ and $\gamma_2 = 0.01$ as the margins of the two loss functions. Here, we use eight A4000-16G GPUs for training the whole model. In addition, δ is set to 0.5, d is set to 768, and r is set to 128.

4.3 Overall performance

Table 1 summarizes our results on the Multi-News dataset. All of these scores follow the original papers. The first part of extractive approaches is the Lead-3 baseline and Oracle upper bound, while the second includes other extractive summarization models. We present our models finally at the bottom. In addition, we also give the LEAD [23] baseline (which selects the first three sentences in a document).

Overall, our HyperSiameseNet outperforms all extractive baseline models. Specifically, MatchSum is similar to our HyperSiameseNet, which formulates the extractive summarization task as a text-matching problem (extract-then-match) in the Euclidean space. From the results in Table 1, we can see that our model is significantly better than MatchSum. We believe this is due to the modeling of text representation in the hyperbolic space and the introduction of the squared poincaré distance as the similarity measurement, which implicitly models the latent hierarchical information between candidate summaries and the document.

4.4 Ablation study

To show the characteristics of HyperSiameseNet and justify the superiority of the hyperbolic space for extractive summarization, we are interested in comparing it with an analogous model in the euclidean space (under different distance measures). The analogous

Model	R-1	R-2	R-L	R-AVG.
LEAD [23]	43.08	14.27	38.97	32.11
MATCH-ORACLE[29]	47.45	17.41	43.14	36.00
TextRank† [19]	41.95	13.86	38.07	31.29
LexRank† [8]	41.77	13.81	37.87	31.15
MMR† [2]	44.72	14.92	40.07	33.24
PG† [16]	44.55	15.54	40.75	33.61
BottomUp† [11]	45.27	15.32	41.38	33.99
Hi-MAP† [9]	45.21	16.29	41.39	34.30
HDSG† [27]	46.05	16.35	42.08	34.83
PRESUMM [28]	46.34	16.88	42.20	35.14
MatchSum [29]	46.20	16.51	41.89	34.87
HyperSiameseNet	**46.67***	**16.93***	**42.38***	**35.33**

Table 1: Performance on Multi-News test set. The models with † indicates that the results provided by [27]. The best results are highlighted in bold. The significance levels (* 0.01) between MatchSum [29] and our model are also provided.

models in euclidean space have a similar architecture as HyperSiameseNet. It directly takes embeddings, which obtains from the pre-trained language model (BERT) as initial word representations in euclidean space. Then, we use the same operation to obtain summary and document representations. Finally, the textual similarities are computed as the distance in euclidean space (e.g., Euclidean distance, Manhattan distance, and Dot product) between their embeddings. The same architecture of the prediction layers is adopted. Concretely, we can see from the results in Table 2 that squared poincaré distance and poincaré distance are better than the distance measures in euclidean space, which benefits from the natural sensitivity of hyperbolic space to hierarchy. It also demonstrates the effectiveness of extractive summarization in the hyperbolic space.

4.5 Sensitivity of Hyper-Parameter

As with the Euclidean space, the dimension of representing embeddings on hyperbolic space is often a key factor affecting model performance. Therefore, we set the dimension size as a hyper-parameter and present the results in a graph.

As can be seen from Figure 3, high dimension does not necessarily represent good results for the hyperbolic spaces. Conversely, lower ranks may lead to better performance. Our model achieves the best results when the dimension is 128.

5 RELATED WORK

Extractive Multi-Document Summarization. Extractive multi-document summarization approaches attempt to generate the final summary by selecting a set of salient sentences from the source document that are most informative and relevant. To tackle this challenging task, various techniques and methods have been applied in extractive summarizers so far [2, 9, 27, 28]. Recent studies have attempted to build two-stage document summarization systems [29]. The first stage is usually to extract some fragments of the original text, and the second stage is to select or modify based on these fragments. Different from the existing extractive summarization models, we explore extractive multi-document summarization in the hyperbolic space instead of the euclidean space.

Model	R-1	R-2	R-L
MatchSum (Cosine Similarity)	46.20	16.51	41.89
MatchSum (Euclidean Distance)	45.78	16.24	41.54
MatchSum (Manhattan Distance)	45.37	16.04	41.15
MatchSum (Dot Product)	45.98	16.42	41.66
HyperSiameseNet			
- Poincaré Distance	46.29	16.65	41.98
- Squared Poincaré Distance	**46.67**	**16.93**	**42.38**

Table 2: Results of different distance measurement on Multi-News test set. The best results are highlighted in bold.

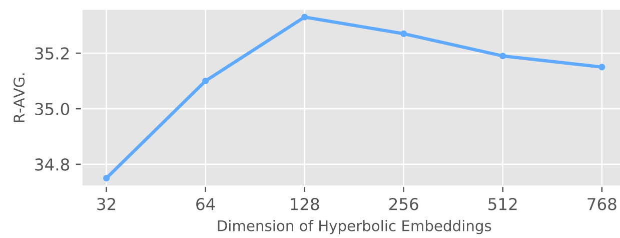

Figure 3: Results of the different hyperbolic embedding ranks of HyperSiameseNet on the Multi-News dataset.

Hyperbolic deep learning. Recent researches have shown that many types of complex data exhibit non-Euclidean structures [10, 13]. The hyperbolic embedding methods have been proposed to learn the latent representation of hierarchical data and demonstrated encouraging results [4, 5, 15, 21, 22]. In the field of natural language processing, hyperbolic representation learning has been successfully applied to generating word embeddings [3, 6, 25]. To our best knowledge, this is the first work to design a hyperbolic matching network for the extractive summarization task.

6 CONCLUSION AND FUTURE WORK

This paper proposes a new hyperbolic siamese network (HyperSiameseNet), which learns representation and models relevance in the same hyperbolic space for extractive summarization. This is the first attempt to design a new hyperbolic siamese network for extractive summarization to the best of our knowledge. In our experiments, HyperSiameseNet outperforms recent state-of-the-art matching-based extractive summarization baselines.

In future work, it would be interesting to incorporate external knowledge (e.g., WordNet) to explicitly assist in learning representations and modeling the latent hierarchical relationships between candidate summaries and their corresponding documents in the hyperbolic space.

7 ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2020AAA0106800; the National Science Foundation of China under Grant 61822601 and 61773050; the Beijing Natural Science Foundation under Grant Z180006; the Fundamental Research Funds for the Central Universities (2019JBZ110).

REFERENCES

[1] Gary Bécigneul and Octavian-Eugen Ganea. 2019. Riemannian Adaptive Optimization Methods. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=r1eiqi09K7

[2] Jaime G. Carbonell and Jade Goldstein. 2017. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *SIGIR Forum* 51, 2 (2017), 209–210. https://doi.org/10.1145/3130348.3130369

[3] Boli Chen, Yao Fu, Guangwei Xu, Pengjun Xie, Chuanqi Tan, Mosha Chen, and Liping Jing. 2021. Probing BERT in Hyperbolic Spaces. In *International Conference on Learning Representations*. https://openreview.net/forum?id=17VnwXYZyhH

[4] Boli Chen, Xin Huang, Lin Xiao, Zixin Cai, and Liping Jing. 2020. Hyperbolic Interaction Model for Hierarchical Multi-Label Classification.. In *AAAI*. AAAI Press, 7496–7503. http://dblp.uni-trier.de/db/conf/aaai/aaai2020.html#ChenHXCJ20

[5] Boli Chen, Xin Huang, Lin Xiao, and Liping Jing. 2020. Hyperbolic Capsule Networks for Multi-Label Classification.. In *ACL*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 3115–3124. http://dblp.uni-trier.de/db/conf/acl/acl2020.html#ChenHXJ20

[6] Shuyang Dai, Zhe Gan, Yu Cheng, Chenyang Tao, Lawrence Carin, and Jingjing Liu. 2020. APo-VAE: Text Generation in Hyperbolic Space. *CoRR* abs/2005.00054 (2020). http://dblp.uni-trier.de/db/journals/corr/corr2005.html#abs-2005-00054

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.. In *NAACL-HLT (1)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. http://dblp.uni-trier.de/db/conf/naacl/naacl2019-1.html#DevlinCLT19

[8] Günes Erkan and Dragomir R. Radev. 2011. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *CoRR* abs/1109.2128 (2011). arXiv:1109.2128 http://arxiv.org/abs/1109.2128

[9] Alexander R. Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir R. Radev. 2019. Multi-News: A Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 1074–1084. https://doi.org/10.18653/v1/p19-1102

[10] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic Neural Networks.. In *NeurIPS*, Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (Eds.). 5350–5360. http://dblp.uni-trier.de/db/conf/nips/nips2018.html#GaneaBH18

[11] Sebastian Gehrmann, Yuntian Deng, and Alexander M. Rush. 2018. Bottom-Up Abstractive Summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 4098–4109. https://doi.org/10.18653/v1/d18-1443

[12] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter W. Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. 2019. Hyperbolic Attention Networks.. In *ICLR (Poster)*. OpenReview.net. http://dblp.uni-trier.de/db/conf/iclr/iclr2019.html#GulcehreDMRPHBB19

[13] Matthias Hamann. 2018. On the tree-likeness of hyperbolic spaces. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 164. 345–361.

[14] C. Hopper and B. Andrews. 2011. *The Ricci Flow in Riemannian Geometry*. The Ricci flow in Riemannian geometry.

[15] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan V. Oseledets, and Victor S. Lempitsky. 2020. Hyperbolic Image Embeddings.. In *CVPR*. IEEE, 6417–6427. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2020.html#KhrulkovMUOL20

[16] Logan Lebanoff, Kaiqiang Song, and Fei Liu. 2018. Adapting the Neural Encoder-Decoder Framework from Single to Multi-Document Summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 4131–4141. https://doi.org/10.18653/v1/d18-1446

[17] C. Y. Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS)*. Barcelona, Spain.

[18] Yang Liu. 2019. Fine-tune BERT for Extractive Summarization. *CoRR* abs/1903.10318 (2019). http://dblp.uni-trier.de/db/journals/corr/corr1903.html#abs-1903-10318

[19] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Text.. In *EMNLP*. ACL, 404–411. http://dblp.uni-trier.de/db/conf/emnlp/emnlp2004.html#MihalceaT04

[20] Maximilian Nickel and Douwe Kiela. 2017. Poincaré Embeddings for Learning Hierarchical Representations.. In *NIPS*. 6338–6347. http://dblp.uni-trier.de/db/conf/nips/nips2017.html#NickelK17

[21] Christopher De Sa, Albert Gu, Christopher Ré, and Frederic Sala. 2018. Representation Tradeoffs for Hyperbolic Embeddings. *CoRR* abs/1804.03329 (2018). arXiv:1804.03329 http://arxiv.org/abs/1804.03329

[22] Rik Sarkar. 2011. Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane.. In *Graph Drawing (Lecture Notes in Computer Science, Vol. 7034)*, Marc J. van Kreveld and Bettina Speckmann (Eds.). Springer, 355–366. http://dblp.uni-trier.de/db/conf/gd/gd2011.html#Sarkar11

[23] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks.. In *ACL (1)*, Regina Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, 1073–1083. http://dblp.uni-trier.de/db/conf/acl/acl2017-1.html#SeeLM17

[24] Mingyang Song, Yi Feng, and Liping Jing. 2022. Hyperbolic Relevance Matching for Neural Keyphrase Extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz (Eds.). Association for Computational Linguistics, 5710–5720. https://doi.org/10.18653/v1/2022.naacl-main.419

[25] Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. 2019. Poincare Glove: Hyperbolic Word Embeddings.. In *ICLR (Poster)*. OpenReview.net. http://dblp.uni-trier.de/db/conf/iclr/iclr2019.html#TifreaBG19

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need.. In *NIPS*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 5998–6008. http://dblp.uni-trier.de/db/conf/nips/nips2017.html#VaswaniSPUJGKP17

[27] Danqing Wang, Pengfei Liu, Yining Zheng, Xipeng Qiu, and Xuanjing Huang. 2020. Heterogeneous Graph Neural Networks for Extractive Document Summarization. *CoRR* abs/2004.12393 (2020). arXiv:2004.12393 https://arxiv.org/abs/2004.12393

[28] Chao Zhao, Tenghao Huang, Somnath Basu Roy Chowdhury, Muthu Kumar Chandrasekaran, Kathleen R. McKeown, and Snigdha Chaturvedi. 2022. Read Top News First: A Document Reordering Approach for Multi-Document News Summarization. *CoRR* abs/2203.10254 (2022). https://doi.org/10.48550/arXiv.2203.10254 arXiv:2203.10254

[29] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2020. Extractive Summarization as Text Matching.. In *ACL*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 6197–6208. http://dblp.uni-trier.de/db/conf/acl/acl2020.html#ZhongLCWQH20