

應用對話語篇剖析於兩階段會議摘要之研究 Leveraging Dialogue Discourse Parsing in a Two-Stage Framework for Meeting Summarization

黃怡萍 Yi-Ping Huang, 羅天宏 Tien-Hong Lo, 陳柏琳 Berlin Chen

國立台灣師範大學資訊工程學系
Department of Computer Science and Information Engineering,
National Taiwan Normal University

yiping97huang@gmail.com, {teinhonglo, berlin}@ntnu.edu.tw

摘要

會議摘要旨在從冗長的會議紀錄中，生成出簡潔並包含重要資訊的文本內容，能夠幫助參與者快速掌握會議的核心要點。然而，會議記錄通常具有複雜的對話結構，如不完整的句子和分散在各個話語中的資訊。此外，文本的長度經常超出了預訓練語言模型能夠處理的長度。本文提出了一種針對「長輸入文本」和「對話式結構」的兩階段摘要生成框架，首先進行文本擷取，從中篩選出重要的文本片段，然後基於這些片段進行摘要生成。對於複雜的對話結構，對話語篇剖析能夠理解話語間的關係，並將其畫成樹狀結構。我們選取較具結構的文本作為擷取階段的輸出，以增加資訊的密度且提供更結構化的對話文本作為生成器的輸入。實驗結果表明，我們的方法可以提升最終生成摘要的表現。

Abstract

Meeting summarization aims to distill meaningful information from lengthy meeting transcripts into concise texts, allowing participants to grasp key points quickly. However, meeting transcripts often feature complex dialogue structures, such as incomplete sentences and information scattered across multiple utterances. Additionally, the length of these transcripts often exceeds the maximum input limit for pretrained language models. In this paper, we introduce a two-stage summarization framework specifically

designed for long-input texts and complex dialogue structures. First, we extract key segments from the original transcript. Second, we generate the summary based on these extracted segments. To address the complexity of dialogue structures, we employ dialogue discourse parsing to comprehend the relationships between utterances, which we represent in a tree-like structure. We select more structured text as the output from the extraction phase to enhance information density, thereby providing a more organized input for the summary generator. Experimental results demonstrate that our approach significantly improves the quality of the generated summaries.

關鍵字：會議摘要、自動文件摘要、對話語篇剖析、生成式模型

Keywords: Meeting Summarization, Automatic Document Summarization, Dialogue Discourse Parsing, Generative Model

1 緒論 (Introduction)

會議是群體和組織間進行討論、決策和交流的主要方式，已成為現代人日常工作中不可或缺的一部分。隨著線上會議的普及和語音辨識技術的快速發展，越來越多的會議文本被轉錄出來，有效的會議摘要能夠提高會議的效率，幫助參與者快速掌握會議的要點，並做出相應的決策。然而，由於會議通常涉及多個參與者，且討論的內容常常包含多個主題，因此要生成有結構、有意義的摘要是一個具有挑戰性的任務。

摘要任務可以分為擷取式 (Extractive) 和重寫式 (Abstractive) 兩種方法，擷取式摘要從原始文本中直接選取一些具有代表性的句子或短語，然後把它們連接起來成為摘要。重寫式摘要則需要對原始文件理解後，重新改寫句子，生成出一個簡潔且包含原始文件中核心內容的摘要。早期的摘要研究主要聚焦於擷取式摘要，因為它相對簡單且不容易產生不正確的資訊。但隨著基於 Transformer (Vaswani et al., 2017) 的序列到序列模型的出現，預訓練語言模型在摘要生成任務上有了重大突破，這也促使研究者逐漸轉向重寫式摘要的研究。然而，在會議摘要方面，存在一些與傳統文本摘要不同的挑戰。Renard et al. (2023) 對這些挑戰進行了整理，包括訓練資料不足或模型限制，以及多方對話中存在的互動特性。訓練資料不足源於會議通常涉及到公司或群體之間的機密，導致公開的會議資料集較少。而模型的限制則是會議的長度往往超越預訓練語言模型能夠處理的長度，這也大幅地增加了實作上的困難。此外，會議是對話的形式，若只是順序建模而未能運用到對話之間的互動關係，也可能導致效果不佳。

在處理超出預訓練語言模型能力範疇的長度時，先前研究提出的解決策略主要可分為四種 (Mao et al., 2022)。首先，採用稀疏注意力 (Sparse Attention) 機制可以降低自注意力機制的複雜度 (Beltagy et al., 2020; Child et al., 2019)，將輸入序列劃分為多個子集，僅對這些子集進行注意力計算。第二種策略是先擷取後生成 (Extract-then-Generate)，先從輸入文本中擷取重要或關鍵資訊，然後使用生成模型進行摘要生成 (Xu and Durrett, 2019; Zhang et al., 2019; Zhang et al., 2021; Zhong et al., 2021)。通過預先篩選，能夠提供生成器更為精簡且重要的資訊。第三種策略是分而治之 (Divide-and-Conquer)，即對每個片段進行摘要生成，再將這些摘要合併以形成整體會議摘要 (Gidiotis and Tsoumakas, 2020; Grail et al., 2021; Zhang et al., 2022)。最後，第四種策略是使用階層式模型 (Hierarchical Model)，對話語的不同結構進行建模，以優化摘要模型 (Rohde et al., 2021; Zhu et al., 2020)。由於先擷取後生成

策略最為直覺且與人類摘要過程相似，本研究採用此方法作為模型的架構。

另一方面，對於多人間的互動，引入外部資訊是一種常見的做法，其中包括 Ganesh and Dingliwal (2020) 透過對話語篇剖析將對話重組成文件，再進行文件摘要。DDAMS (Feng et al., 2021) 將帶有語篇關係的會議話語轉化為會議圖，然後運用圖編碼器對其進行建模。Goo and Chen (2018) 則利用對話行為為摘要模型提供了更多資訊，並設計了句子門控機制來建模對話行為和摘要之間的關係。在這些外部資訊中，對話語篇剖析能夠提供話語之間的預定義關係 (Chi and Rudnicky, 2022)，明確標示出話語之間的資訊流和互動，進而建構出整個會議的結構。

基於上述，本研究針對「長輸入文本」和「對話式結構」提出了先擷取後生成的模型框架，在擷取階段透過對話語篇剖析來選擇文本片段中較具有結構的文本內容，既能降低文本的冗餘度和增加資訊密度，又可提供更精簡且結構化的對話文本作為生成器的輸入。生成階段透過先進的生成式預訓練模型 DialogLED (Zhong et al., 2022) 來生成最終摘要。

2 相關研究 (Related Work)

2.1 先擷取後生成架構 (Extract-then-Generate Framework)

之前的先擷取後生成模型可以分為三個類別。第一類方法是分別訓練擷取器 (Extractor) 和生成器 (Generator) (Lebanoff et al., 2019; Xu and Durrett, 2019; Zhang et al., 2019; Zhang et al., 2021)。然而，這種方法在將擷取的片段傳遞給生成器時會有資訊損失的問題。為了解決這一問題，後來的研究使用強化學習來連接這兩個階段以達到改進效果，例如 Chen and Bansal (2018) 和 Bae et al. (2019) 的研究，它們分別採用句子級別和摘要級別的策略梯度。

不過，當前的強化學習方法主要使用句子級別的 ROUGE 和摘要級別的 ROUGE 作為訓練獎勵。在這其中，使用句子級別的 ROUGE 可能會選擇具有重疊內容的句子，導致生成摘要的冗餘度較高。另外，使用摘要級別的 ROUGE 會導致訓練訊號的稀疏性，而更長的

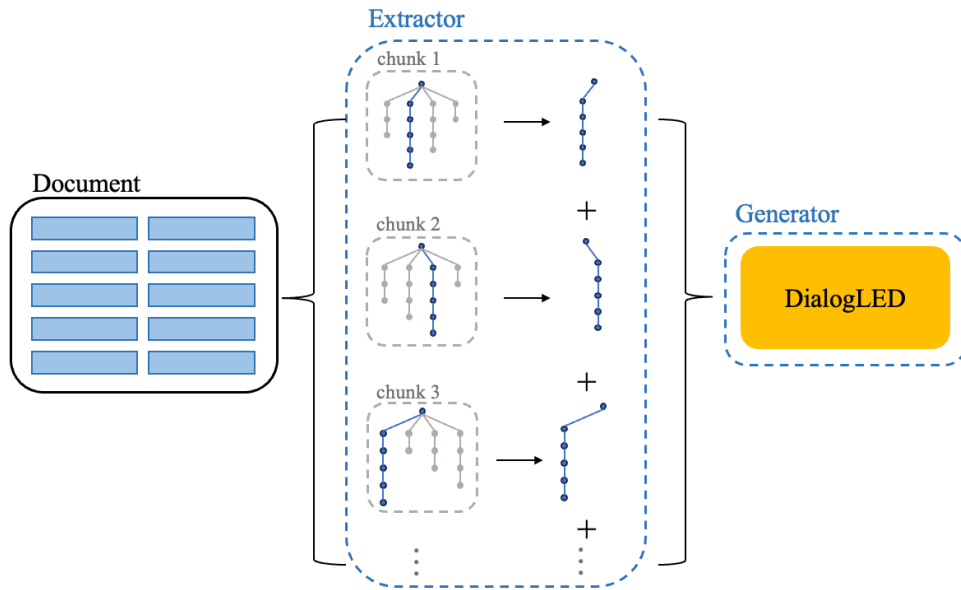


圖 1. 模型架構圖。首先，會將會議文本以句子為單位，平均切分成 10 等份的文本區塊 (Chunk)。擷取器接著將每個區塊透過對話語篇剖析建構成語篇圖，並從中選取最長鏈作為輸出。最後將每個區塊的輸出結合起來，傳遞到生成器 DialogLED 做摘要生成。

輸入使得這種方法變得更加困難。為了克服這些困難，Mao et al. (2022) 提出了一種聯合訓練的方法。在這個方法中，生成器會在每個時間步驟中為每個擷取的片段動態地分配權重。平均動態權重可以表示每個擷取片段的重要性，通過這種方式可以增強擷取器的訓練效果。這樣可以同時訓練擷取器和生成器，從而更好地兼顧擷取和生成兩個階段。

2.2 引入對話語篇剖析 (Incorporating Dialogue Discourse Parsing)

在自然語言處理中，語篇結構已被利用於語言生成和機器閱讀理解等任務 (Joty et al., 2019)。Gerani et al. (2014) 和 Louis et al. (2010) 在文本摘要中引入語篇關係，並證明了方法的有效性。

根據 Rennard et al. (2023) 的研究，Feng et al. (2021) 是第一個將對話語篇剖析技術應用於會議資料集 AMI (Carletta et al., 2005) 和 ICSI (Janin et al., 2003)，進行生成式摘要的任務。在這項研究中，他們使用了一個順序性語篇剖析器 (Afantenos et al., 2015)，在對話式的語料 STAC (Asher et al., 2016) 上訓練後，接著透過這個模型，他們能夠獲取會議資料的語篇關係圖 (Discourse Relation Graph)。將會議轉

換成語篇圖後，透過 Pointer-Generator Network (PGN) (See et al., 2017) 解碼器來生成最終摘要。Ganesh and Dingliwal (2020) 提出了一種兩步驟、無監督的重寫式會議摘要模型，首先利用話語關係將會議重構成類似文件的文本，然後使用主流的預訓練文件摘要模型 BART (Lewis et al., 2020) 生成摘要。

Chen and Yang (2021) 則是使用話語圖為對話資料集 SAMSum (Gliwa et al., 2019) 生成短對話摘要。還增加了動作圖 (Action Graph) 來提取話語中的三元組 *who - doing - what* 結構，然後通過圖注意網路 (Graph Attention Network) (Veličković et al., 2018) 對話語和動作圖進行編碼，最後將它們與解碼器 BART 連接來生成摘要。

3 方法 (Method)

3.1 任務定義 (Task Definition)

給定一個會議文件 $D = \{u_1, \dots, u_L\}$ ，包含 L 個話語輪次 (Turn)。接著，跟一般擷取式摘要不同的是，會先把輸入文件分成多個區塊 (Chunk)，而區塊的大小考量到生成模型 DialogLED 的預訓練，其一次參考總文本長度 10% 的會議文本作為窗口，從中做隨機遮罩

Noise Type	Description	Example
Speaker Mask	Randomly mask 50% of the speakers.	[MASK]: Good morning! How are you today?
Turn Splitting	Split a single turn into multiple turns. Keep the speaker of the first turn and mask the rest.	Tom: Good morning! [MASK]: How are you today?
Turn Merging	Merge multiple turns into one turn. Keep the first speaker and remove the rest.	Tom: Good morning! How are you today? I'm doing well, thank you. How about you?
Text Infilling	Mask the content of the dialogue.	Tom: Good morning! How are you [MASK]?
Turn Permutation	Shuffle the order of the turns within the dialogue.	Bob: I'm doing well, thank you. How about you? Tom: Good morning! How are you today?

表 1. 對話相關的雜訊。包含語者遮罩 (Speaker Mask)、話語輪次分割 (Turn Splitting)、話語輪次合併 (Turn Merging)、文本填充 (Text Infilling) 和話語輪次交換 (Turn Permutation)。模型在預訓練時，透過這些機制學習對話相關的知識。

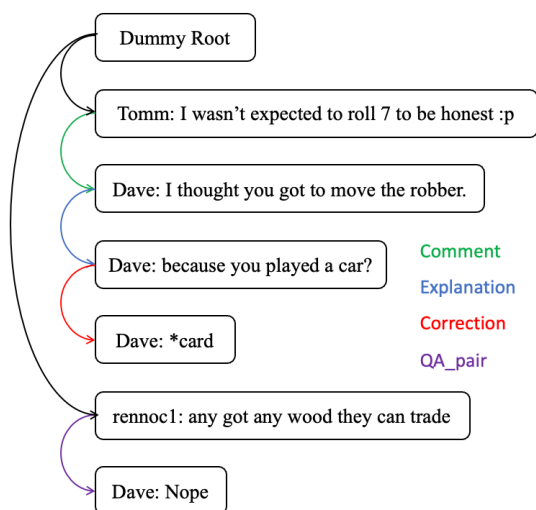


圖 2. 對話語篇剖析示意圖

和預測。因此，這裡我們也將區塊的大小訂為 10% 的輸入長度。經過句子級別的切分後，每個會議文本會有 10 個句數相近的區塊 $C = \{c_0, \dots, c_9\}$ ，每個區塊的句子數為 $l = \lfloor \frac{L}{10} \rfloor$ 。

在擷取階段，我們把每個區塊透過對話語篇剖析畫成一個生成樹 (Spanning Tree)，根據長度排序會得到 $\{R_1, R_2, \dots\}$ 個鏈。我們探討了 $Top-1 = \{R_1\}$ ， $Top-2 = \{R_1, R_2\}$ 及 $Top-3 = \{R_1, R_2, R_3\}$ 三種選擇對於結果的影響。

接著，用 $C' = \{c'_0, \dots, c'_9\}$ 表示經過篩選後的文本區塊，作為生成模型的輸入。在生成階段的目標是要生成一個長度為 T 的摘要 $Z = \{z_1, z_2, \dots, z_T\}$ ，給定輸入 C' 和先前生成的文字 $z_{<t}$ ，計算 $P(Z|C') = \prod_{t=1}^T P(z_t|C', z_{<t})$ 。

3.2 模型架構 (Model Structure)

模型的架構如圖 1 所示，首先將輸入文本切成多個區塊，擷取階段會在每個文本區塊選擇 Top-1 至 Top-3 個最長鏈，再將選出的鏈結組合起來作為生成器的輸入。生成階段使用的模型是訓練在對話語料上的生成式預訓練語言模型 DialogLED。

3.3 擷取器 (Extractor)

對話語篇剖析旨在通過找出所有話語鏈結和對應的關係，來揭示多參與者對話的內部結構。圖 2 為模型在 STAC 資料集上所預測出的話語關係示意圖。

本研究使用的對話語篇剖析模型 SDDP (Chi and Rudnicky, 2022) 是目前的 State-of-the-art 模型。在編碼步驟透過兩個 Bi-LSTM 將兩兩話語對和對應的鏈結關係建構成三維的向量空間表示。在解碼步驟使用最大生成樹演算法，得到最終的話語關係樹。

	Max	Mean	Min
Baseline	10269	4338.74	572
Top-1	3425	1686.43	350
Top-2	4093	2514.58	407
Top-3	4876	3088.58	492

表 2. 輸入長度分析

	Rouge-1	Rouge-2	Rouge-L
Baseline	52.34	18.45	49.76
Top-1	52.70	19.66	50.61
Top-2	51.61	18.25	49.27
Top-3	52.58	18.65	50.36

表 3. 最終實驗結果的 ROUGE 分數

Models	Rouge-1	Rouge-2	Rouge-L
BART-large ($\ell = 3072$)	51.77	18.83	49.67
Longformer ($\ell = 8192$)	54.20	20.72	51.36
UniLM-base ($\ell = 5120$)	51.92	18.42	49.89
DialogLM ($\ell = 5120$)	54.49	20.03	51.92
DialogLED ($\ell = 5120$)	54.80	20.37	52.26
Baseline ($\ell = 2560$)	52.34	18.45	49.76
Top-1 ($\ell = 2560$)	52.70	19.66	50.61

表 4. 實驗結果與基準模型比較。其中 Baseline 為 DialogLED 的 2560 tokens 版本，Top-1 是透過我們提出的方法訓練 Baseline 模型。

3.4 生成器 (Generator)

DialogLED 是一個基於窗口去噪的預訓練語言模型，其模型架構為 Longformer-Encoder-Decoder (LED) (Beltagy et al., 2020)，並使用設計的對話相關去噪任務來預訓練 LED 模型。對話相關雜訊如表 1 所示，共設計了 5 種雜訊，分別和語者、話語內容和話語間相對順序有關。

給定一個包含 n 個話語輪次的對話文本 $D = (x_1, x_2, \dots, x_n)$ ，話語輪次表示一個語者 - 話語對 (Speaker-Utterance Pair) $x_i = (s_i, u_i)$ 。接著，隨機選擇一個包含多個輪次的窗口 $W = (x_j, x_{j+1}, \dots, x_{j+m})$ ，將對話相關的雜訊加入，成為新的帶雜訊窗口 $\tilde{W} = (\tilde{x}_j, \tilde{x}_{j+1}, \dots, \tilde{x}_{j+m})$ 。

在預訓練階段，用帶雜訊的窗口取代原本的並與其他所有話語輪次連接成一個長序列，作為模型的輸入。解碼器要能夠透過帶雜訊的窗口和對話其餘部分來重建成原始窗口。

下游任務中，LED 的編碼器讀取文件，解碼器生成輸出摘要。其編碼器不使用完整的自注意力，而是使用 Longformer 的稀疏注意力。解碼器對整個編碼的文本和先前解碼的位置使用完整的自注意力機制。

4 實驗分析 (Experimental Analysis)

4.1 資料集 (Dataset)

本研究使用的資料集 AMI (Carletta et al., 2005) 是目前最廣泛應用的英語會議語料庫。此語料庫收集了一個設計團隊的會議內容，其中有四名參與者分別為：專案經理 (Project Manager, PM)、行銷專家 (Marketing Expert, ME)、使用者介面設計師 (User Interface Designer, UI) 和工業設計師 (Industrial Designer, ID)，共同討論設計和開發一個新的電視遙控器。

4.2 實驗設置 (Settings)

我們的生成模型使用 Huggingface 上開源的 DialogLED-large-5120¹，但由於記憶體限制，

¹ <https://huggingface.co/MingZhong/DialogLED-large-5120>

	Precision	Recall	F1
Baseline	0.861	0.855	0.858
Top-1	0.866	0.858	0.862
Top-2	0.863	0.856	0.860
Top-3	0.863	0.856	0.860

表 5. 最終實驗結果的 BERTScore 分數

我們將最大輸入長度由 5120 tokens 調整為 2560 tokens。在輸入時，超過最大長度的文本將會被截斷。表 2 分析了資料集中最大長度、平均長度和最小長度，Baseline 為資料集的原始長度，Top-1、Top-2 和 Top-3 為經過我們的方法篩選後的長度。生成摘要的最大和最小長度設定為 512 tokens 和 256 tokens。

4.3 評估指標 (Evaluation Metrics)

我們使用的評估指標：

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004) 是評估自動文件摘要常用的方法。以召回率 (Recall) 作為核心評估指標，用於衡量自動生成的摘要與參考摘要之間的字詞相似程度。

BERTScore (Zhang et al., 2020) 用於評估文本生成任務 (例如文件摘要和機器翻譯) 的輸出與真實參考之間的相似性。它基於 BERT 的上下文嵌入做計算，能夠捕捉更深層的語義和語境資訊。

4.4 基準模型 (Baseline Models)

我們將結果與以下幾個強基準模型做比較：

BART (Lewis et al., 2020)：是目前在短文本生成中，最先進的去噪序列到序列預訓練模型。透過特殊的去噪目標來預訓練模型，使其能夠在後續的微調任務中，如文本摘要和文本翻譯等，取得卓越的效果。

Longformer (Beltagy et al., 2020)：針對長序列處理的預訓練模型。設計了局部與全局注意機制，能夠在保持計算效率的同時，處理比標準 Transformer 模型更長的文本。

UniLM (Dong et al., 2019)：一種統一的序列到序列預訓練模型，其設計目的是為了在多個 NLP 任務上，如文本生成、文本翻譯和文本摘要等，取得好的表現。

DialogLM (Zhong et al., 2022)：在 UniLM 上引入對話相關任務的預訓練模型。通過專門針對對話資料的預訓練，來捕捉對話之間的上下文關係和語言模式，從而在對話相關的下流任務上取得更好的效果。

DialogLED (Zhong et al., 2022)：將與 DialogLM 同樣的對話相關任務應用於 Longformer 模型中，不僅可以處理長對話，還優化了對話之間的語境和關係。

4.5 實驗結果 (Experimental Results)

表 3 為最終實驗結果的 ROUGE 分數，透過我們提出的方法，在擷取階段選擇較具結構的文本並過濾掉分散的話語，能夠為生成結果帶來提升，其中，僅選擇最長鏈的效果最為顯著。

表 4 與其他基準模型做比較，在輸入長度最短的情況下，所提出方法的表現能夠優於部分基準模型，但仍然低於最先進的模型 DialogLED-large-5120。

我們另外透過 BERTScore 來比較我們的方法與原始輸入的差異。表 5 表明了經過我們的方法篩選出的文本能夠使生成模型生成出語義上更接近參考摘要的結果。

5 結論 (Conclusion)

本研究提出了一個適用於會議摘要的兩階段先擷取後生成摘要模型，針對於會議摘要任務中的「長輸入文本」和「對話式結構」兩個限制分別提出改進的方法。透過先擷取後生成架構，能夠解決生成模型無法一次讀取過長序列的問題。同時，引入對話語篇剖析來篩選文本片段有助於提供生成模型資訊更為集中且更具結構的輸入文本，並過濾掉對話中的冗餘內容。實驗結果表明，我們的方法能夠為摘要結果帶來提升，並且生成出語義更為接近的摘要。

在未來的研究中，我們期望能整合更多樣的外部知識到我們的摘要模型，如對話行為和對話語篇剖析的關係類別，以及包括聲音和影像在內的多模態資訊。此外，大型語言模型 (Large Language Model, LLM) 是未來研究的趨勢，所以如何精準地微調及運用這些大型模型也將是我們關注的重點。

References

- Stergos Afantenos, Eric Kow, Nicholas Asher, and J  r  my Perret. 2015. Discourse parsing for multi-party chat dialogues. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 928–937, Lisbon, Portugal. Association for Computational Linguistics.
- Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2721–2727, Portoro , Slovenia. European Language Resources Association (ELRA).
- Sanghwan Bae, Taek Kim, Jihoon Kim, and Sang-goo Lee. 2019. Summary Level Training of Sentence Rewriting for Abstractive Summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 10–20, Hong Kong, China. Association for Computational Linguistics.
- Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. arXiv:2004.05150.
- Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. 2005. The AMI meeting corpus: a pre-announcement. In *Proceedings of the Second international conference on Machine Learning for Multimodal Interaction*, pages 28–39, Berlin, Heidelberg. Springer-Verlag.
- Jiaao Chen and Diyi Yang. 2021. Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1380–1391, Online. Association for Computational Linguistics.
- Yen-Chun Chen and Mohit Bansal. 2018. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 675–686, Melbourne, Australia. Association for Computational Linguistics.
- Ta-Chung Chi and Alexander Rudnicky. 2022. Structured Dialogue Discourse Parsing. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 325–335, Edinburgh, UK. Association for Computational Linguistics.
- Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019. Generating Long Sequences with Sparse Transformers. arXiv:1904.10509.
- Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified Language Model Pre-training for Natural Language Understanding and Generation. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- Xiachong Feng, Xiaocheng Feng, Bing Qin, and Xinwei Geng. 2021. Dialogue Discourse-Aware Graph Model and Data Augmentation for Meeting Summarization. In volume 4, pages 3808–3814.
- Prakhar Ganesh and Saket Dingliwal. 2020. Restructuring Conversations using Discourse Relations for Zero-shot Abstractive Dialogue Summarization. arXiv:1902.01615.
- Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bitan Nejat. 2014. Abstractive Summarization of Product Reviews Using Discourse Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1602–1613, Doha, Qatar. Association for Computational Linguistics.
- Alexios Gidiotis and Grigorios Tsoumakas. 2020. A Divide-and-Conquer Approach to the Summarization of Long Documents. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 28:3029–3040.
- Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. SAMSUM Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 70–79, Hong Kong, China. Association for Computational Linguistics.
- Chih-Wen Goo and Yun-Nung Chen. 2018. Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized by Dialogue Acts. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 735–742.
- Quentin Grail, Julien Perez, and Eric Gaussier. 2021. Globalizing BERT-based Transformer Architectures for Long Document Summarization. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1792–1810, Online. Association for Computational Linguistics.
- Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. 2003. The ICSI Meeting

- Corpus. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, volume 1.
- Shafiq Joty, Giuseppe Carenini, Raymond Ng, and Gabriel Murray. 2019. Discourse Analysis and Its Applications. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 12–17, Florence, Italy. Association for Computational Linguistics.
- Logan Lebanoff, Kaiqiang Song, Franck Dernoncourt, Doo Soon Kim, Seokhwan Kim, Walter Chang, and Fei Liu. 2019. Scoring Sentence Singletons and Pairs for Abstractive Summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2175–2189, Florence, Italy. Association for Computational Linguistics.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Annie Louis, Aravind Joshi, and Ani Nenkova. 2010. Discourse indicators for content selection in summarization. In *Proceedings of the SIGDIAL 2010 Conference*, pages 147–156, Tokyo, Japan. Association for Computational Linguistics.
- Ziming Mao, Chen Henry Wu, Ansong Ni, Yusen Zhang, Rui Zhang, Tao Yu, Budhaditya Deb, Chenguang Zhu, Ahmed Awadallah, and Dragomir Radev. 2022. DYLE: Dynamic Latent Extraction for Abstractive Long-Input Summarization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1687–1698, Dublin, Ireland. Association for Computational Linguistics.
- Virgile Rennard, Guokan Shang, Julie Hunter, and Michalis Vazirgiannis. 2023. Abstractive Meeting Summarization: A Survey. arXiv:2208.04163.
- Tobias Rohde, Xiaoxia Wu, and Yinhan Liu. 2021. Hierarchical Learning for Generation with Long Source Sequences. arXiv:2104.07545.
- Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. arXiv:1704.04368.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. arXiv:1710.10903.
- Jiacheng Xu and Greg Durrett. 2019. Neural Extractive Text Summarization with Syntactic Compression. 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)*, pages 3292–3303, Hong Kong, China. Association for Computational Linguistics.
- Haoyu Zhang, Jingjing Cai, Jianjun Xu, and Ji Wang. 2019. Pretraining-Based Natural Language Generation for Text Summarization. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 789–797, Hong Kong, China. Association for Computational Linguistics.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. arXiv:1904.09675.
- Yusen Zhang, Ansong Ni, Ziming Mao, Chen Henry Wu, Chenguang Zhu, Budhaditya Deb, Ahmed Awadallah, Dragomir Radev, and Rui Zhang. 2022. Summ^N: A Multi-Stage Summarization Framework for Long Input Dialogues and Documents. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1592–1604, Dublin, Ireland. Association for Computational Linguistics.
- Yusen Zhang, Ansong Ni, Tao Yu, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, and Dragomir Radev. 2021. An Exploratory Study on Long Dialogue Summarization: What Works and What’s Next. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4426–4433, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2022. DialogLM: Pre-trained Model for Long Dialogue Understanding and Summarization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):11765–11773.
- Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, and

Dragomir Radev. 2021. QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5905–5921, Online. Association for Computational Linguistics.

Chenguang Zhu, Ruochen Xu, Michael Zeng, and Xuedong Huang. 2020. A Hierarchical Network for Abstractive Meeting Summarization with Cross-Domain Pretraining. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 194–203, Online. Association for Computational Linguistics.