ABSTRACTIVE DIALOGUE SUMMARIZATION WITH SENTENCE-GATED MODELING OPTIMIZED BY DIALOGUE ACTS

Chih-Wen Goo[†⋆] *and Yun-Nung Chen*[†]

[†]National Taiwan University, Taipei, Taiwan
[⋆]MediaTek Inc., Hsinchu, Taiwan
r05944049@ntu.edu.tw y.v.chen@ieee.org

ABSTRACT

Neural abstractive summarization has been increasingly studied, where the prior work mainly focused on summarizing single-speaker documents (news, scientific publications, etc). In dialogues, there are diverse interactive patterns between speakers, which are usually defined as dialogue acts. The interactive signals may provide informative cues for better summarizing dialogues. This paper proposes to explicitly leverage dialogue acts in a neural summarization model, where a sentence-gated mechanism is designed for modeling the relationships between dialogue acts and the summary. The experiments show that our proposed model significantly improves the abstractive summarization performance compared to the state-of-the-art baselines on the AMI meeting corpus, demonstrating the usefulness of the interactive signal provided by dialogue acts.[1]

Index Terms— dialogues, summarization, dialogue act, sentence gate, gating mechanism.

1. INTRODUCTION

With a large amount of textual information available, text summarization has been widely studied for several years in natural language processing, which can be categorized into two types: *extractive summarization* and *abstractive summarization*. Extractive methods assemble the summary from the source text directly [1, 2, 3], while abstractive methods generate words to form the summary [4, 5, 6]. With the rising trend of neural models, abstractive summarization has been widely investigated recently [7, 5]. In addition, some recent work proposed to combine advantages from two types of methods and achieved better summarization results [8, 9, 10, 11].

Most of the summarization work focused on single-speaker written documents such as news, scientific publications, etc [12, 7, 13]. In addition to text summarization, speech summarization is equally important especially for spoken or even multimedia documents, which are more difficult to browse than text, such as multi-party meetings. Therefore, speech summarization has been investigated in the past [14, 15, 16, 17, 18, 19]. However, almost all prior work focused on summarizing the documents based on the mentioned salient content instead of the interactive status, but this behavioral signal should be important for dialogue summarization.

To better summarize a meeting, not only the content but also the inter-speaker interactions are important. Prior dialogue summarization work utilized prosody or speaker information as interactive patterns for better extracting salient sentences [18, 20]. However, abstractive summarization for dialogue/meeting summarization has not yet explored due to the lack of suitable benchmark data [21], because the benchmark dialogue data is only annotated with the importance of utterances without abstractive summaries [22]. In order to bridge the gap, this paper benchmarks the abstractive dialogue summarization task using the AMI meeting corpus [22], where the summaries are produced based on the annotated topics the speakers discuss about. A *topic* or a *high-level description* of a meeting is treated as the abstractive summary; for instance, "*evaluation of project new idea for TV*" is a summary of the meeting topics. Such dialogue summaries are very short and may not contain words directly mentioned by the speakers, making automatic summarization more challenging.

A dialogue is a sequence of utterances interacting between multiple participants, where each utterance would modify both participants' cognitive status and the current dialogue state. The effect of an utterance on the context is often called a *dialogue act* [23], which provides informative cues for better understanding dialogues. Therefore, dialogue act classification has been widely studied in the spoken language understanding research field, and previous work about dialogue act recognition used information sources from multiple modalities, including linguistic information, global contextual properties like knowledge about participants, and so on [24, 25, 26, 27, 28]. Popular approaches for dialogue act classification include support vector machine (SVM) [29], Naive Bayes [26, 30, 31], logistic regression [32], and recur-

This work is done while the author was at National Taiwan University.

[1]The source code is available at http://github.com/MiuLab/DialSum.

Multi-Party Dialogue	Dialogue Act
A: mm-hmm .	Backchannel
B: mm-hmm .	Backchannel
C: then , these are some of the remotes which are different in shape and colour , but they have many buttons .	Inform
C: so uh sometimes the user finds it very difficult to recognise which button is for what function and all that .	Inform
D: so you can design an interface which is very simple , and which is user-friendly .	Inform
D: even a kid can use that .	Inform
A: so can you got on t t uh to the next slide .	Suggest
Summary: alternative interface options	

Fig. 1. A dialogue instance in the dataset built from the AMI meeting corpus.

rent neural network (RNN) [33, 34, 35, 36].

Dialogue act classification and summarization are usually treated independently and used for different goals. In this paper, we leverage dialogue act information to improve dialogue summarization. Assuming that dialogue acts, indicating interactive signals, may be important for better summarization, how to effectively integrate the information into a neural summarization model is the main focus of this paper. Prior work attempted at modeling the discourse information and proposed a discourse-aware summarization model using the hierarchical RNN [37], where the between-utterance cues are modeled in an implicit way. Also, they performed the model in a publication summarization task, where the input documents are relatively structured, and there is no interactive behavior in such documents.

Therefore, this work focuses on how to effectively model the interactive signals such as dialogue acts for better dialogue summarization, where we introduce a sentence-gated mechanism to jointly model the explicit relationships between dialogue acts and summaries. To the best of our knowledge, there is no previous study with the similar idea, and we summarize our contributions as three-fold:

- The proposed model is the first attempt for dialogue summarization using dialogue acts as explicit interactive signals.
- We benchmark the dataset for abstractive summarization in the meeting domain, where the summaries describe the high-level goals of meetings.
- Our proposed model achieves the state-of-the-art performance in dialogue summarization and helps us analyze how much each utterance and its dialogue act affect the summaries.

2. DIALOGUE SUMMARIZATION DATASET

Considering that there is no abstractive summarization data in any conversational domain, this paper first builds a dataset in order to benchmark the experiments. The AMI meeting corpus is a well-known meeting data with different annotations [22], which consists of 100 hours of meeting recordings. The recordings use a range of signals synchronized to a common timeline, including close-talking and far-field microphones, individual and room-view video cameras, and output from a slide projector and an electronic whiteboard. The meetings are recorded in English using three different rooms with different acoustic properties, and include mostly non-native speakers. It contains a wide range of annotations, including dialogue acts, topic descriptions, named entities, hand gesture, and gaze direction. In this work, we use the recording transcripts as the input to our model. Because there is no summary annotation in the AMI data, the annotated topic descriptions are treated as summaries of the dialogues. In AMI data, the annotations for dialogue acts and topic descriptions are not available for all utterances, so we extract a subset of the AMI corpus to construct the benchmark dialogue summarization dataset. Figure 1 is an example dialogue instance, where the summary describes the high-level goal of the meeting.

AMI Corpus	Statistics
Vocabulary Size	8,886
#Dialogue Act	15
Min Summary Length	1
Max Summary Length	26
Training Set Size	7,024
Development Set Size	400
Testing Set Size	400

Table 1. Statistics of the AMI meeting corpus for dialogue summarization.

We use a sliding window size of 50 words to split a meeting into several dialogue samples, where we adjust the boundary to make sure no utterance would be cut in the middle. If the topic changes within the window, all topic descriptions are concatenated according to their appearing order. In each resulting sample, there are around 50 to 100 words in an arbitrary number of sentences. We extract 7,824 samples from 36 meeting recordings and then randomly separate them into three groups: 7,024 samples for training, 400 samples for development, and 400 samples for testing. There are 15 dialogue act labels in the training set. The detailed statistics are shown in Table 1.

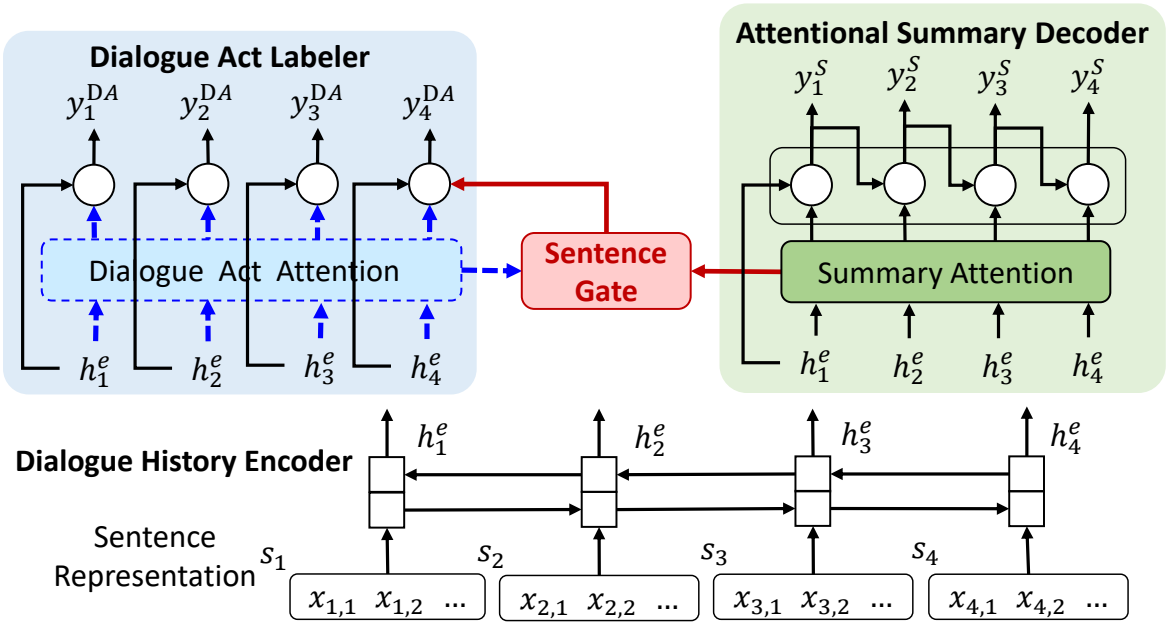

Fig. 2. The architecture of the proposed sentence-gated models.

3. PROPOSED APPROACH

This section first explains our attention-based RNN model and then introduces the proposed sentence gating mechanism for summarization jointly optimized with dialogue act recognition. The model architecture is illustrated in Figure 2, where there are several modules including 1) a *dialogue history encoder*, 2) a *dialogue act labeler*, 3) an *attentional summary decoder*, and 4) a *sentence gate*. We detail each module below.

3.1. Dialogue History Encoder

Given a dialogue document, there is a sequence of utterances $\mathbf{s} = (s_1, \ldots, s_K)$ as the input, where K is the dialogue length. An utterance is constituted by a word sequence $\mathbf{x} = (x_1, \ldots, x_T)$, and the sentence embedding can be obtained by averaging all word embeddings in that sentence[2]. The bidirectional long short-term memory (BLSTM) model [38] takes a sentence sequence $\mathbf{s}$ as the input, and then generates a forward hidden state $\overrightarrow{h_i^e}$ and a backward hidden state $\overleftarrow{h_i^e}$. The final hidden state h_i^e at the time step i is the concatenation of $\overrightarrow{h_i^e}$ and $\overleftarrow{h_i^e}$, i.e. $h_i^e = [\overrightarrow{h_i^e}, \overleftarrow{h_i^e}]$, which can be viewed as the encoded information for the given source document.

[2]The experiments using RNN-learned sentence embeddings are conducted, but the performance is similar to using the average of word embeddings. Considering the parameter size, all experiments use average vectors as sentence embeddings

3.2. Dialogue Act Labeler

To leverage the dialogue act information, this module focuses on predicting dialogue acts for all utterances. Specifically, $\mathbf{s}$ is mapping to its corresponding dialogue act label $\mathbf{y} = (y_1^{DA}, \ldots, y_K^{DA})$. For each hidden state h_i, we compute the dialogue act context vector c_i^{DA} as the weighted sum of LSTM's hidden states, $h_1^e, ..., h_T^e$, by the learned attention weights $\alpha_{i,j}^{DA}$:

$$c_i^{DA} = \sum_{j=1}^{K} \alpha_{i,j}^{DA} \cdot h_j^e, \tag{1}$$

where the dialogue act attention weights are computed as

$$\alpha_{i,j}^{DA} = \frac{\exp(e_{i,j})}{\sum_{k=1}^{K} \exp(e_{i,k})}, \tag{2}$$

$$e_{i,k} = \sigma(W_{he}^{DA} \cdot h_k^e), \tag{3}$$

where σ is the $\texttt{sigmoid}$ activation function, and W_{he}^{DA} is the weight matrix of a feed-forward neural network. Then all hidden states and dialogue act context vectors are optimized for dialogue act modeling by

$$y_i^{DA} = \texttt{softmax}(W_{hy}^{DA} \cdot (h_i^e + c_i^{DA})), \tag{4}$$

where y_i^{DA} is the dialogue act label of the i-th sentence in the given dialogue, and W_{hy}^{DA} is the weight matrix. The dialogue act attention is shown as the blue component in Figure 2.

3.3. Attentional Summary Decoder

Following the prior work [10, 37], we use an attentional decoder for generating the word sequence as the summary. The

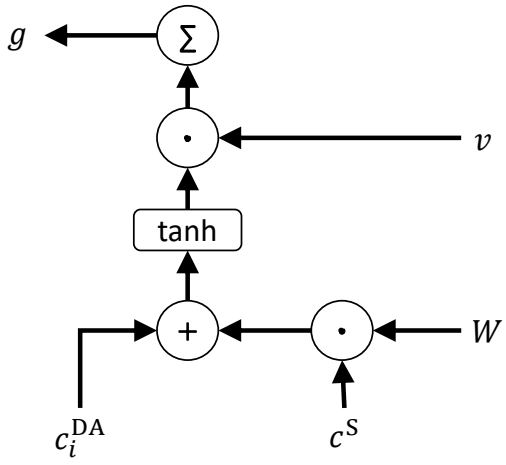

Fig. 3. Illustration of the sentence gate.

summary context vector c_i^S is computed as c^{DA} similarly:

$$c_i^S = \sum_{j=1}^{K} \alpha_{i,j}^S \cdot h_j^e. \tag{5}$$

The summary is generated by a unidirectional LSTM with the initial state set to be h_K^e, the last hidden state of the dialogue history encoder. The unidirectional LSTM will output words until generating an end-of-string token or reaching the predefined maximum length. The formulation is shown as:

$$y_i^S = \texttt{softmax}(W_{hy}^S \cdot (h_i^d + c_i^S)). \tag{6}$$

3.4. Sentence-Gated Mechanism

A gating mechanism is able to model the explicit relationship between two types of information [39]. The proposed sentence-gated model introduces an additional gate that leverages a summary context vector for modeling relationships between dialogue acts and summaries in order to improve the dialogue act labeler and the attentional summary decoder illustrated in Figure 3. The proposed model has two different types:

- Full attention
  The model considers the relations from dialogue acts and summaries using both *dialogue act attention* and *summary attention* shown as the blue and green blocks respectively in Figure 2.
- Summary attention
  The model builds the gating mechanism using only *summary attention*, where the parameter size is smaller than the full attention model.

3.4.1. Full Attention

First, a dialogue act context vector c_i^{DA} and an averaged summary context vector c^S are combined to pass through a slot gate:

$$c^S = \frac{1}{K} \sum_{k=1}^{K} c_i^S, \tag{7}$$

$$g = \sum v \cdot \tanh(c_i^{DA} + W \cdot c^S), \tag{8}$$

where v and W are a trainable vector and a matrix respectively. The summation is done over elements in one time step. g can be seen as a weighted feature of the joint context vector (c_i^{DA} and c^S). We use g to weight between h_i and c_i^{DA} for deriving y_i^{DA} and then replace (4) as below:

$$y_i^{DA} = \texttt{softmax}(W_{hy}^{DA} \cdot (h_i + c_i^{DA} \cdot g)). \tag{9}$$

A larger g indicates that the dialogue act context vector and the summary context vector pay attention to the similar part of the input sequence, which also infers that the correlation between the dialogue act and the summary is stronger and the context vector is more *reliable* for contributing the prediction results.

3.4.2. Summary Attention

To deeply investigate the power of the sentence gate mechanism, we eliminate the dialogue act attention module in the architecture, so c_i^{DA} is replaced with h_i^e. Accordingly, (8) and (9) are reformed as (10) and (11) respectively,

$$g = \sum v \cdot \tanh(h_i^e + W \cdot c^S) \tag{10}$$

$$y_i^{DA} = \texttt{softmax}(W_{hy}^{DA} \cdot (h_i + h_i \cdot g)) \tag{11}$$

This version allows the dialogue acts and summaries to share the attention mechanism, so both information would be mutually improved in a more direct manner compared to the full attention version.

3.5. Joint End-to-End Training

To learn the summarization model optimized by the dialogue act information, we formulate a joint objective as

$$\begin{aligned} p(y^{DA}, y^S \mid \mathbf{s}) & \\ &= \prod_{k=1}^{K} p(y_t^S \mid s_k) \prod_{k=1}^{K} p(y_t^{DA} \mid s_k) \\ &= \prod_{k=1}^{K} p(y_t^S \mid \mathbf{x}_k) \prod_{k=1}^{K} p(y_t^{DA} \mid \mathbf{x}_k), \end{aligned} \tag{12}$$

where $p(y^{DA}, y^S \mid \mathbf{s})$ is the conditional probability of dialogue acts and the summary given the input dialogue. Based on the joint objective, the proposed model that utilizes interactive signals for summarization can be trained in an end-to-end fashion.

Model		Interactive Signal	Size	DA	Summarization			
				Acc	R-1	R-2	R-3	R-L
BLSTM Dialogue Act Labeler		△	3,864K	64.16	—	—	—	—
Attentional Seq2Seq [6]		✗	12,391K	—	34.74	25.15	21.35	34.70
Pointer-Generator Network [10]		✗	11,861K	—	31.21	26.35	25.22	31.21
Discourse-Aware Hierarchical Seq2Seq [37]		△	11,295K	—	66.82	37.74	27.71	47.84
Proposed	Sentence-Gated (Full Attention)	✓	12,363K	**64.47**†	67.52†	37.38	27.70	48.45†
	Sentence-Gated (Summary)	✓	11,837K	64.28	**68.34**†	**39.25**†	**29.05**†	**49.93**†

Table 2. Performance on the AMI meeting data (%). † indicates the significant improvement over all baselines with $p < 0.05$.

4. EXPERIMENTS

To evaluate the proposed model, we conduct experiments using the AMI meeting data introduced in Section 2.

4.1. Setup

In all experiments, the optimizer is `adam`, the reported numbers are averaged over 20 runs, and the maximum epoch is set to 30 with an early-stop strategy. In our proposed model, the size of hidden vectors are set to 256, and the vector dimensions vary for the compared baselines such that all models have the similar size.

For evaluation metrics, the dialogue act performance is measured by the accuracy (Acc), and the summary performance is measured by ROUGE-1 (R-1), ROUGE-2 (R-2), ROUGE-3 (R-3), and ROUGE-L (R-L) scores [40]. We also validate the performance improvement with a statistical significance test for all experiments, where single-tailed t-test is performed to measure whether the results from the proposed model are significantly better than all baselines. The dag symbols indicate the significant improvement with $p < 0.05$.

4.2. Baselines

Considering that there is no previous work for joint dialogue act modeling and summarization, the compared baselines are either for dialogue act classification or text summarization, including a bidirectional LSTM for dialogue act labeler, an attentional seq2seq summarization model [6], a pointer-generator network [10], and a discourse-aware hierarchical attentional seq2seq [37]. Please note that the BLSTM dialogue act labeler baseline is the same as our proposed model without the summarization component. The pointer-generator network extends the attentional seq2seq by adding a joint pointer network to enable the copy mechanism, For the discourse-aware model, we only use the concept about the hierarchy introduced by Cohan et al. [37] but do not include its pointer network part. The reason will be latter explained in Section 4.3. Among all baselines, only the discourse-aware model implicitly utilizes the interactive signal, while our model explicitly optimizes the summary together with dialogue acts.

4.3. Results

The experimental results are shown in Table 2, where the models have similar size of parameters. Among all summarization baselines, the discourse-aware hierarchical seq2seq model achieves better performance than other two baselines, indicating the importance of discourse/interaction cues for dialogue summarization. Comparing between attentional seq2seq and the pointer-generator network, the difference is not obvious, because the high-level descriptions as summaries do not overlap between the input dialogues and the corresponding summaries (1.2% of the overlapping rate for AMI meeting data). Therefore, due to the low overlapping rate, the pointer-generator network performs the worst, because the pointer network and coverage loss parts introduce noises. This is the reason that other baselines and our proposed model do not contain the copy mechanism and coverage loss in the experiments. The finding suggests that the dialogue summarization focuses more on the *interaction goal* instead of the *mentioned content*.

Table 2 shows that the proposed sentence-gated mechanism with summary attention significantly outperforms all baselines, where almost all measurements obtain the significant improvement, demonstrating that interactive signal provides useful cues for dialogue summarization, and the proposed sentence-gated mechanism effectively models the relationships between them. The proposed model with full attention performs slightly worse than the one with summary attention only. The probable reason is that the dialogue act attention may not be necessary for predicting the dialogue acts of a single utterance; that is, dialogue acts are often decided only based on the individual utterance, so adding attention on its contextual utterances may not bring much benefit for modeling such interactive behaviors. Moreover, the proposed model reduces the model size by 12% compared to the best baseline combination (BLSTM for dialogue act prediction + discourse-aware hierarchical attention seq2seq for summarization) and demonstrates the better model capacity.

4.4. Attention Analysis

To further analyze the attention learned in the model, we visualize the utterance attention weights when generating sum-

Testing Dialogue Example 1	Dialogue Act
A: okay .	Assess
B: okay that's fine , that's good .	Assess
C: okay , let's start from the beginning	Offer
C: so i'm going to speak about technical functions design	Inform
C: un just like some some first issues that came up .	Inform
B: um 'kay ,	Stall
C: so the method i was um adopting at this point , it's not um for the for the whole um period of the um all the project but it's just at th at this very moment .	Inform
B: um	Stall
C: uh my method was um to look at um other um remote controls ,	Inform
C: uh so mostly just by searching on the web	Inform
C: and to see what um functionality they used .	Inform
C: and then um after having got this inspiration and having compared what i found on the web um just to think about what the de what the user really needs and what um what the use might desire as additional uh functionalities .	Inform
Generated summary: industrial designer presentation issues of participants	
Reference summary: industrial designer presentation interface specialist presentation	

Testing Dialogue Example 2	Dialogue Act
A: okay , so	Stall
B: hmm , okay .	Backchannel
A: yeah well uh	Stall
A: ipod is trendy .	Inform
A: and it is well curved square .	Inform
C: yeah .	Backchannel
A: square . like .	Inform
B: yeah , but mm is uh has round corners i think .	Assess
A: so	Stall
D: we shouldn't have too square corners and that kind of thing .	Inform
Generated summary: look and usability	
Reference summary: look and usability	

Fig. 4. Visualization of summary attention vectors. The darker color indicates higher attention wights. The underlined word is the target word for illustrating the attention.

maries in Figure 4. Figures are colored with different levels of the summary attention, where the darker one has a larger attention value as its importance when generating the target word, and vice versa. It is obvious that the proposed model successfully captures which ones are the key sentences in the dialogues. It may be credited to the proposed sentence gate that learns the dialogue acts conditioned on its summary in order to provide the helpful signal for global optimization of the joint model. In addition, it can be found that the "Inform" dialogue act usually guides the model to pay more attention to it, which aligns well with our intuition. In sum, for dialogue summarization, the experiments show that modeling dialogue acts and summary relations controlled by the novel sentence-gated mechanism can effectively improve abstractive summarization performance in terms of ROUGE scores due to the joint optimization with dialogue act modeling.

5. CONCLUSION

This paper focuses on abstractive dialogue summarization by modeling interactive behaviors, where the proposed model uses a novel sentence-gate that allows the dialogue act signal can be conditioned on the learned summarization result, in order to achieve better performance for both tasks. This paper benchmarks the experiments using a meeting dataset, and the experiments show that the proposed approach outperforms all state-of-the-art models, demonstrating the importance of interactive cues in dialogue summarization.

6. ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their insightful feedback on this work. The authors are financially supported by Ministry of Science and Technology (MOST) in Taiwan and MediaTek Inc.

7. REFERENCES

[1] Julian Kupiec, Jan Pedersen, and Francine Chen, "A trainable document summarizer," in *Proceedings of SIGIR*, 1995, pp. 68–73.

[2] Yun-Nung Chen, Yu Huang, Ching-Feng Yeh, and Lin-Shan Lee, "Spoken lecture summarization by random walk over a graph constructed with automatically extracted key terms," in *Proceedings of INTERSPEECH*, 2011, pp. 933–936.

[3] Horacio Saggion and Thierry Poibeau, "Automatic Text Summarization: Past, Present and Future," in *Multi-source, Multilingual Information Extraction and Summarization*, R. Yangarber T. Poibeau; H. Saggion. J. Piskorski, Ed., Theory and Applications of Natural Language Processing, pp. 3–13. Springer, 2012.

[4] Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A Smith, "Toward abstractive summarization using semantic representations," in *Proceedings of NAACL-HLT*, pp. 1077–1086.

[5] Sumit Chopra, Michael Auli, and Alexander M Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in *Proceedings of NAACL-HLT*, 2016, pp. 93–98.

[6] Ramesh Nallapati, Bing Xiang, and Bowen Zhou, "Sequence-to-sequence rnns for text summarization," *CoRR*, 2016.

[7] Alexander M Rush, Sumit Chopra, and Jason Weston, "A neural attention model for abstractive sentence summarization," in *Proceedings of EMNLP*, 2015, pp. 379–389.

[8] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li, "Incorporating copying mechanism in sequence-to-sequence learning," in *Proceedings of ACL*, Berlin, Germany, 2016, pp. 1631–1640.

[9] Yishu Miao and Phil Blunsom, "Language as a latent variable: Discrete generative models for sentence compression," in *Proceedings of EMNLP*, 2016, pp. 319–328.

[10] Abigail See, Peter J Liu, and Christopher D Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of ACL*, 2017, vol. 1, pp. 1073–1083.

[11] Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, Jing Tang, and Min Sun, "A unified model for extractive and abstractive summarization using inconsistency loss," in *Proceedings of ACL*, 2018, pp. 1–10.

[12] Hung-yi Lee, Sz-Rung Shiang, Ching-feng Yeh, Yun-Nung Chen, Yu Huang, Sheng-Yi Kong, and Lin-shan Lee, "Spoken knowledge organization by semantic structuring and a prototype course lecture system for personalized learning," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 5, pp. 883–898, 2014.

[13] Sebastian Gehrmann, Yuntian Deng, and Alexander M Rush, "Bottom-up abstractive summarization," in *Proceedings of EMNLP*, 2018.

[14] Sameer Maskey and Julia Hirschberg, "Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization," in *Proceedings of EUROSPEECH*, 2005.

[15] David Harwath and Timothy J Hazen, "Topic identification based extrinsic evaluation of summarization techniques applied to conversational speech," in *Proceedings of ICASSP*, 2012, pp. 5073–5076.

[16] Korbinian Riedhammer, Benoit Favre, and Dilek Hakkani-Tür, "Long story short–global unsupervised models for keyphrase based meeting summarization," *Speech Communication*, vol. 52, no. 10, pp. 801–815, 2010.

[17] Yun-Nung Chen, "Automatic key term extraction and summarization from spoken course lectures," M.S. thesis, National Taiwan University, 6 2011.

[18] Yun-Nung Chen and Florian Metze, "Two-layer mutually reinforced random walk for improved multi-party meeting summarization," in *Proceedings of SLT*, 2012, pp. 461–466.

[19] Yun-Nung Chen and Florian Metze, "Multi-layer mutually reinforced random walk with hidden parameters for improved multi-party meeting summarization," in *Proceedings of INTERSPEECH*, 2013, pp. 485–489.

[20] Yun-Nung Chen and Florian Metze, "Intra-speaker topic modeling for improved multi-party meeting summarization with integrated random walk," in *Proceedings of NAACL-HLT*, 2012, pp. 377–381.

[21] Mahak Gambhir and Vishal Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017.

[22] Iain McCowan, Jean Carletta, W Kraaij, S Ashby, S Bourban, M Flynn, M Guillemot, T Hain, J Kadlec, V Karaiskos, et al., "The ami meeting corpus," in *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*, 2005, vol. 88, p. 100.

[23] Harry Bunt, "Context and dialogue control," *THINK Quarterly*, vol. 3, 1994.

[24] Ken Samuel, Sandra Carberry, and K. Vijay-Shanker, "Dialogue act tagging with transformation-based learning," in *Proceedings of COLING*, 1998.

[25] Helen Wright, Massimo Poesio, and Stephen Isard, "Using high level dialogue information for dialogue act recognition using prosodic features," in *ESCA Tutorial and Research Workshop (ETRW) on Dialogue and Prosody*, 1999.

[26] Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Computational Linguistics*, vol. 26, no. 3, pp. 339–373, Sept. 2000.

[27] Tina Klüwer, Hans Uszkoreit, and Feiyu Xu, "Using syntactic and semantic based relations for dialogue act recognition," in *Proceedings of COLING*, 2010, pp. 570–578.

[28] Quan Hung Tran, Ingrid Zukerman, and Gholamreza Haffari, "Preserving distributional information in dialogue act classification," in *Proceedings of EMNLP*, 2017, pp. 2151–2156.

[29] Maryam Tavafi, Yashar Mehdad, Shafiq Joty, Giuseppe Carenini, and Raymond Ng, "Dialogue act recognition in synchronous and asynchronous conversations," in *Proceedings of SIGDIAL*, 2013, pp. 117–121.

[30] Simon Keizer, Rieks op den Akker, and Anton Nijholt, "Dialogue act recognition with bayesian networks for dutch dialogues," in *Proceedings of SIGDIAL*, 2002, pp. 88–94.

[31] J. Ang, Yang Liu, and E. Shriberg, "Automatic dialog act segmentation and classification in multiparty meetings," in *Proceedings of ICASSP*, 2005, pp. 1061–1064.

[32] Yun-Nung Chen, William Yang Wang, and Alexander I Rudnicky, "An empirical investigation of sparse log-linear models for improved dialogue act classification," in *Proceedings of ICASSP*. IEEE, 2013, pp. 8317–8321.

[33] Yangfeng Ji, Gholamreza Haffari, and Jacob Eisenstein, "A latent variable recurrent neural network for discourse-driven language models," in *Proceedings of NAACL-HLT*, 2016, pp. 332–342.

[34] Hamed Khanpour, Nishitha Guntakandla, and Rodney D. Nielsen, "Dialogue act classification in domain-independent conversations using a deep recurrent neural network," in *Proceedings of COLING*, 2016.

[35] Ji Young Lee and Franck Dernoncourt, "Sequential short-text classification with recurrent and convolutional neural networks," in *Proceedings of NAACL-HLT*, 2016, pp. 515–520.

[36] Nal Kalchbrenner and Phil Blunsom, "Recurrent convolutional neural networks for discourse compositionality," in *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, 2013, pp. 119–126.

[37] Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian, "A discourse-aware attention model for abstractive summarization of long documents," in *Proceedings of NAACL-HLT*, 2018, vol. 2, pp. 615–621.

[38] Grégoire Mesnil, Yann Dauphin, Kaisheng Yao, Yoshua Bengio, Li Deng, Dilek Hakkani-Tur, Xiaodong He, Larry Heck, Gokhan Tur, Dong Yu, et al., "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 3, pp. 530–539, 2015.

[39] Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen, "Slot-gated modeling for joint slot filling and intent prediction," in *Proceedings of NAACL-HLT*, 2018, pp. 753–757.

[40] Chin-Yew Lin, "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, 2004.