Documentation and Evaluation of Gated Fusion for Dialogue Act Classification

Abstract_ *This document presents a gated fusion mechanism for dialogue act classification that integrates speaker and topic metadata into a BiLSTM encoder via learnable gates. Evaluations on SWDA, MRDA, and DYDA corpora demonstrate consistent improvements of 2–4% in test accuracy over baseline models without metadata. Detailed experiments, ablation studies, and error analyses provide insights into the contributions of each gating component and outline directions for further enhancement.*

I. Introduction:

A. Background and Motivation:

Dialogue act classification is a core task in conversational AI, aiming to assign communicative functions (e.g., question, statement, acknowledgement) to utterances. Recent advances in transformer-based architectures have shown strong performance; however, exploiting metadata such as speaker identity and conversation topic remains underexplored. Gated fusion mechanisms offer a principled approach to integrate auxiliary information by learning gating weights that modulate contributions of different feature streams.

B. Objectives:

This document presents the design, implementation, and evaluation of gated fusion models for dialogue act classification on three corpora: Switchboard Dialogue Act (SWDA), Meeting Recorder Dialogue Act (MRDA), and Daily Dialogue Act (DYDA). We aim to:

1. Describe model architectures and training pipelines.
2. Detail experimental setups and hyperparameter choices.
3. Present quantitative results and analyze the impact of gated fusion.
4. Conduct an error analysis and discuss limitations.
5. Provide recommendations for future work.

II. Methodology:

A. Datasets:

1. SWDA:

- Contains 1,155 telephone conversations annotated with 42 dialogue act labels.

- Preprocessing includes tokenization, speaker turn segmentation, and label mapping.
-

2. MRDA:

- Multi-party meeting corpus with 5 broad dialogue act categories.
- Each utterance involves multiple speakers; speaker embedding contextualizes turn-taking.

3. DYDA:

- Daily conversations dataset annotated with 4 dialogue act classes.
- Includes topic metadata enabling topic-conditioned models.

B. Model Architecture:

1. Base_encoder:

- 4-layer BiLSTM with hidden size 256, dropout 0.5.
- Word embeddings initialized from pretrained GloVe (300d) and fine-tuned during training.

2. Auxiliary Encoders:

- **Speaker Encoder**: embedding lookup for speaker IDs (size 32);
- **Topic Encoder**: embedding lookup for topics (size 64) when available.

3. Gated Fusion Layer:

- Gating mechanisms compute scalar gates and for speaker and topic streams: $g_s = \sigma(W_s[h;\ s]\ +\ b_S),\ g_t = \sigma(W_t[h;\ t]\ +\ b_t)$
- Fused representation: $h' = h + g_s \odot s + g_t \odot t$

4. Classification Head:

Two fully connected layers with ReLU activations, followed by softmax over nclass.

C. Training Procedure:

- Adam optimizer, learning rate schedules depending on corpus (e.g., 2e-5 for SWDA).
- Early stopping on validation accuracy with patience 5.

- Batch sizes tuned per dataset (e.g., 8 for SWDA, 4 for MRDA, 10 for DYDA).
- Mixed GPU training on 2 GPUs with data parallelism.

III. Experimental Setup:

A. Hyperparameters:

Corpus	batch_size	emb_batch	epochs	lr	layers	chunk_size	dropout	speaker_info	topic_info
SWDA	8	0	30	2e-05	2	196	0.5	gated	none
MRDA	4	256	100	1e-04	1	350	0.5	gated	none
DYDA	10	0	100	1e-04	2	0	0.5	gated	emb_cls

B. Evaluation Metrics:

- **Accuracy**: percentage of utterances correctly classified.
- **Confusion Matrix**: to identify systematic errors between classes.
- **Per-Class Precision/Recall/F1**: to assess performance on imbalanced labels.

IV. Results:

A. Quantitative Performance:

Exp ID	Corpus	Val Acc	Test Acc	Train Loss
E04	SWDA	0.847	0.835	0.471
E05	MRDA	0.901	0.925	0.243
E06	DYDA	0.862	0.886	0.332

B. Ablation Study:

- **No Gating**: removal of gated fusion decreases test Acc by 2–3% across corpora.
- **Speaker Only**: gating speaker info yields smaller gains on SWDA (0.5%) than MRDA (1.5%).
- **Topic Only**: topic gating has negligible impact on DYDA when speaker gating also present.

C. Learning Curves:

- Training and validation loss curves indicate stable convergence by 20 epochs.
- Minimal overfitting observed, possibly due to gating regularization effect.

V. Evaluation and Discussion:

A. Comparative Analysis:

- Our gated fusion model outperforms baseline BiLSTM (no metadata) by 3–4% test accuracy.
- Comparable to recent transformer-based approaches but with fewer parameters and lower compute.

B. Error Analysis:

- Confusion between BACKCHANNEL and AGREEMENT in SWDA suggests gating insufficient to capture discourse nuance.
- MRDA model underperforms on ACTION-NEG categories, indicating potential need for external context.
- DYDA shows misclassifications in SPEECH-ACT-3 under low-resource topics.

C. Strengths and Limitations:

- **Strengths**: lightweight gating, improved generalization, modular architecture.
- **Limitations**: reliance on accurate speaker and topic metadata; gating adds minimal overhead but may over-regularize.

D. Future Work:

- Extend gating to incorporate dialogue history vectors.
- Explore hierarchical gating across turns and sessions.
- Integrate pretrained transformer encoders (e.g., BERT) with gating layers.

VI. Conclusion:

This documentation details the design and evaluation of gated fusion mechanisms for dialogue act classification. Empirical results demonstrate consistent accuracy improvements across three distinct conversational corpora. Error analysis highlights avenues for further model enhancements. Future work will investigate richer contextual gating and integration with transformer backbones.

References:

[1] J. Arevalo, T. Solorio, M. Montes-y-Gómez, and F. A. González, “Gated Multimodal Units for Information Fusion,” arXiv preprint arXiv:1702.01992, Feb. 2017.

[2] Z. He, L. Tavabi, K. Lerman, and M. Soleymani, “Speaker Turn Modeling for Dialogue Act Classification,” in *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic, Nov. 2021, pp. 2150–2157, doi: 10.18653/v1/2021.findings-emnlp.185.

Limitations of This Project

While our work advances dialogue act classification through gated speaker and topic information, several limitations remain:

1. **Dataset Scope and Diversity**
   - We evaluated only on three corpora (SWDA, MRDA, DYDA), each with its own annotation style and domain. Results may not generalize to other dialogue genres (e.g., customer service chats, multi-party group conversations).
   - Our models assume relatively clean transcripts,noisy real-world ASR outputs or code-switched dialogues may degrade performance.

2. **Fixed Ontology of Dialogue Acts**
   - We use a fixed set of act labels (43 for SWDA, 5 for MRDA, 4 for DyDA). In open-domain settings, new or nuanced act types may arise that our classifiers cannot capture without retraining and label expansion.

3. **Resource Requirements**
   - Gated speaker/topic modules increase model complexity and GPU memory usage. Smaller teams or industry practitioners with limited compute may struggle to reproduce our exact setups.

4. **Limited Temporal Dynamics**
   - We model speaker turns and local context but do not explicitly encode longer-range temporal dialogue structures (e.g., conversational goals, topic shifts over many turns).

5. **Evaluation Metrics**
   - Our primary metrics (accuracy on validation and test) do not reflect downstream impact (e.g., improvement in dialogue system response quality). Future work should include human assessments or task-oriented evaluations.