論文 / 著書情報
Article / Book Information

題目(和文)	文脈を考慮した対話における感情認識
Title(English)	
著者(和文)	石渡太智
Author(English)	Taichi Ishiwatari
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12914号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:徳永 健伸,岡崎 直観,村田 剛志,宮崎 純,齋藤 豪
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12914号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

博士論文

文脈を考慮した対話における感情認識

2024 年 8 月 15 日

指導教員　徳永 健伸 教授
提出者　東京工業大学
情報理工学院 情報工学系 知能情報コース
石渡 太智

概 要

対話における感情認識は，会話における各発話の感情を認識する技術であり，ソーシャルメディアでの感情分析や感情的かつ共感的な対話システムの構築に利用される重要な技術として注目を集めている．対話の感情認識を実現するためには，話者が示す感情の種類を定義し，定めた感情ラベルを対話の各発話に付与して，作成したデータセットを基に対象発話の感情ラベルを予測する識別モデルを構築する必要がある．また，識別モデルを構築するためには，発話間の関係や会話の履歴，話者固有の特徴，常識的知識などの対話の感情認識タスク特有の課題に取り組む必要がある．

本研究は，対話の感情認識における上記の研究課題の中でも，特に識別モデルの構築に着目し，発話間の関係と会話の履歴を利用する新たな識別モデルを提案する．感情の種類の定義と感情ラベルの付与に対して，本研究は，識別モデルの構築に取り組む従来研究と同様に，Ekman の 6 感情を基にしたベンチマークデータセットを用いる．識別モデルの構築では，課題の解決によって高い認識性能を示すことが期待され，多くの従来研究の注目を集める発話間の関係と会話の履歴に着目する．

まず発話間の関係の課題では，発話の距離を利用する手法を提案する．対話の感情認識では，発話間の関係の中で，自分自身の感情の推移を表す自己依存と他者の発話が影響を与える他者依存の関係が，話者の感情に影響を与えることが知られている．従来手法の多くは，グラフニューラルネットワークを用いて自己依存と他者依存の関係を利用し，高い認識性能を示した．しかしながら，これらの依存関係を利用する手法は，発話の距離を考慮しない課題が存在する．話者の感情は，しばしば直近の発話の影響を受けるなど，発話の距離に依存する．従って，発話の距離は感情認識を行う上で重要な情報である．そこで本研究は，発話間の関係に加えて，対象の発話から周辺の発話への距離の情報も利用する手法を提案する．提案手法を用いることで，自己依存と他者依存を含む発話間の関係と，発話の距離の両方を利用することができる．対話の感情認識における 3 つのベンチマークデータによる評価実験を通して，提案手法の有効性を確認した．

次に，会話の履歴の課題では，発話間の関係を利用する識別モデルと，会話の履歴を利用する識別モデルを組み合わせる手法を提案する．対話の感情認識では，同じ発話であっても，一連の会話の履歴に応じて異なる感情を示すことがある．会話の履歴を利用する代表的な方法として，連続した複数の発話を連結し，言語モデルに入力する方法がある．この手法は，識別対象の発話とその先行文脈に注意を向けるため，一連の会話の履歴を考慮することができる．しかしながら，この手法は，会話全体に注意を向けるため，逆に個々の発話の依存関係の利用が容易でない．

そこで，本研究は発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組み合わせるアンサンブル手法を提案する．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いものを検索し，動的な重み付き線形和によって補強する事例ベース手法を提案する．具体的には，識別対象の発話とその先行文脈をクエリーとして，会話の履歴の観点で意味的に近い発話を訓練データセットからｋ近傍法を用いて検索する．検索した発話（近傍事例）に付与された感情ラベルと，識別対象の発話との距離を基に感情ラベルの確率分布を作成し，発話間の関係を利用するモデルの確率分布と，動的な重み付き線形和によって組み合わせる．提案手法を用いることで，発話間の関係と会話の履歴の両方の特徴を利用することが可能となる．対話の感情認識の３つのベンチマークデータによる評価実験を通して，動的に重み係数を変更する提案手法が，最高水準の認識性能を示し，その有効性を確認した．

本研究の貢献を示す．本研究は，対話の感情認識において，発話と発話の距離の情報を利用するために，発話間の関係を利用するグラフニューラルネットワークに，はじめて距離の情報を付与する方法を提案した．提案手法を用いることで，自己依存と他者依存を含む発話間の関係と，発話の距離の両方の利用を可能にした．従来手法との比較実験を通して，提案手法は従来手法を上回る最高水準の認識性能を示し，その有効性を確認した．また，本研究は，発話間の関係を利用する識別モデルと，会話の履歴を利用する識別モデルを組み合わせる手法を提案した．２つの異なるモデルを組み合わせる方法として，近傍事例を活用し，はじめて対話の感情認識タスクに適用した．単純に組み合わせるだけでなく，識別対象の発話に応じて動的に変化する重み係数を用いて発話間の関係を利用するモデルの確率分布と，近傍事例による確率分布を組み合わせた．従来手法との比較実験を通して，重み係数を動的に変更する提案手法は従来手法を上回る最高水準の認識性能を示し，その有効性を確認した．

目 次

図 目 次

表 目 次

第1章　序論

1.1　対話における感情認識

対話における感情認識 (ERC: Emotion Recognition in Conversations) は，会話における各話者の発話の感情を認識する技術であり，ソーシャルメディアでの感情分析 [Poria et al., 2019] や感情的かつ共感的な対話システムの構築 [Majumder et al., 2020] に利用される重要な技術として注目を集めている．また裁判や面接，ヘルスケアサービスへの応用も期待される重要な技術である [Poria et al., 2019]．

対話の感情認識は，対話の内容と話者の情報から，発話ごとに適切な感情ラベルを識別することを目的とする．例えば，表 1.1 のように，8 つの発話で構成される対話の場合，1 番から 8 番までの対話の内容と話者の情報を基に，各発話の感情ラベル，例えば 4 番目の発話の感情ラベル *frustrated* を識別する．文章や単一の発話を認識する従来の感情認識タスクとは異なり，複数の発話による文脈や話者固有の特徴などの会話特有の性質が話者の感情に影響を与える特徴を持つ．

順番	話者	発話	感情ラベル
1	A	I'm just so tired all the time.	*sad*
2	B	Well have you been trying to get a job, look for a job or...?	*neutral*
3	A	I've been looking for like eight months.	*frustrated*
4	B	I know., It- It's really tough out there., It's really hard to find a job.	*frustrated*
5	A	I'm tired of the same excuses., No, no you're not qualified enough, wish you had more education.	*frustrated*
6	B	Well what are you looking for?, I mean–	*neutral*
7	B	Well, okay. Well that's–	*neutral*
8	A	Cause I went to Harvard.	*anger*

表 1.1: 対話の各発話に付与された感情ラベルの例

1.2　対話の感情認識の研究課題

対話の感情認識を実現するためには，幾つかの研究課題に取り組む必要がある．一連の会話の文脈や話し手の特徴，感情の移り変わりなどのタスク特有の課題が存在し，課題解決に向けてこれまでに多くの研究がなされてきた [Poria et al., 2019, Pereira et al., 2022]．本節では，それぞれの研究課題を説明し，従来研究の取り組みを示す．

1.2.1　感情の種類の選択

各発話の感情は，あらかじめ定めた感情ラベルによって定義される．一般的な感情ラベルに，Plutchik の 8 感情がある [Plutchik, 1982]．Plutchik の感情は，主に 8 つのタイプ (*joy*, *trust*, *fear*, *surprise*, *sadness*, *disgust*, *anger*, *anticipation*) に分類され，さらに各タイプはその感情の強さに応じて関連する感情が対応づけられる．例えば，感情ラベル *fear* は，恐怖の度合いがより強い *terror* と対応づけられる．また，より単純化された感情モデルに，Ekman の 6 感情がある [Ekman, 1993]．Ekman は，(*anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*) の 6 つの基本的な感情を定義した．

感情ラベルとして，Plutchik の感情を選択する場合，話者が表現する複雑な感情を定義することが可能である．しかし，評価者による適切な感情ラベルの選択が容易でないため，評価者間のラベルの一致度が低下する．一方で，単純化された Ekman の 6 感情を選択する場合，評価者間のラベル一致度は向上する．しかし，付与した感情ラベルと話者が示す本来の感情との間に乖離が生じる可能性がある．このように，感情ラベルの表現力と正確さにはトレードオフが存在する．昨今，公開されているベンチマークデータセットのほとんどは，単純かつ直感的な性質を持つ Ekman の 6 感情を基に，僅かな修正を加えた感情ラベルを利用する傾向にある [Busso et al., 2008, Poria et al., 2018, Zahiri and Choi, 2018]．

1.2.2　感情ラベルの付与

次に，定義した感情ラベルを対話の各発話に付与する．会話の話者が自分自身の発言に感情ラベルを付与する方法が最も望ましいが，リアルタイムにラベルを付与することは会話の流れに影響を与えてしまうため実現が不可能である．会話が終了した後に話者にヒアリングすることも可能であるが，そのような方法を採用する研究は未だ存在しない [Poria et al., 2019]．

別のアプローチとして，適切な感情が自然に引き出されるような仮想的な会話のシナリオを用意し，会話の台本や各発話の感情を制御する方法がある [Busso et al., 2008]．この場合，台本に記された感情を感情ラベルとして利用することが可能で

あるが，実際の会話は想定したシナリオと異なる展開を示すことがあり，シナリオの感情を適切に反映しない．

このような理由から，書き起こした会話に事後的に第三者が感情ラベルを付与する方法が一般的である [Busso et al., 2008, Poria et al., 2018, Zahiri and Choi, 2018]．しかしながら，この方法は話者の感情が評価者の視点に依存する課題が存在する [Poria et al., 2019]．いずれのベンチマークデータセットも，話者の感情を適切に反映させるために，会話を書き起こしたスクリプトだけでなく会話の様子を収録した動画も利用し話者の感情を評価する [Busso et al., 2008, Poria et al., 2018, Zahiri and Choi, 2018]．しかし，話者の感情は依然として評価者の視点に依存するため，異なる評価者間のラベルの不一致が起こる．

1.2.3 識別モデルの構築

対話を書き起こし，話者の情報と感情ラベルを付与したデータセットを構築した後は，対話の内容から各発話の感情ラベルを認識する識別モデルを構築する必要がある．文章や単一の発話を認識する従来の感情認識と異なり，発話間の関係や話者固有の状態などの会話特有の性質を考慮する必要がある．以下に，対話の感情認識タスク特有の課題を示す．

発話間の関係

対話の感情認識では，ある話者は他の話者の感情にしばしば影響を与えるため，発話と発話の関係が重要とされている [Poria et al., 2019]．発話間の関係の中でも，自身の発話からの影響 (自己依存) と他者の発話からの影響 (他者依存) が重要である [Ghosal et al., 2019, Shen et al., 2021]．表 1.1 を用いて，2 つの依存関係の重要性を示す．表 1.1 は，2 人の話者が就職活動について意見を交わす例である．話者 A は長い間就職先が見つからないため，一連の発話で常に負の感情を抱いている．これは自己依存の例を示し，自分自身の感情の推移を表す．一方で，話者 B の 4 番目の感情は，直前の話者 A の状況に同情し，負の感情を抱いている．これは他者依存の例を示し，他者の発話が自身の感情に影響を与える性質を持つ．このように，議論の構造や意図，対話者への態度など，様々な要因に左右される発話間の関係は，話者の感情を識別する上で重要である．

これまでに多くの研究が，自己依存と他者依存を含む発話間の関係を利用する方法を提案した [Ghosal et al., 2019, Sheng et al., 2020, Shen et al., 2021]．Ghosal らは，識別対象の発話に対して，周辺のある発話が自分自身によるものか他者によるものか，または過去に発せられたか未来に発せられたかによって依存関係を区別する手法を提案した [Ghosal et al., 2019]．また，Sheng らも同様に，発話単位で自己依存と他者依存を考慮する手法を提案した [Sheng et al., 2020]．最新の研究として，Shen らは自己依存と他者依存の関係を利用するモデリングの方法に着

順番	話者	発話	感情ラベル
1	Monica	That's good, have a seat.	*neutral*
2	Monica	Um, the doctor says it's gotta be a needle.	*neutral*
3	Monica	You're just gonna have to be brave, ok?	*neutral*
4	Monica	Can you do that for me?	*neutral*
5	Ross	Ok.	*neutral*
6	Monica	Ok. Oh boy. You are doin' so good. You wanna squeeze my hand?	*neutral*
7	Ross	**Yes!**	*fear*

表 1.2: 負の感情を示す発話 "Yes" の例

順番	話者	発話	感情ラベル
1	Mark	Rachel?	*neutral*
2	Rachel	Yeah. Hi Mark!	*surprise*
3	Mark	Hi. I just talked to Joanna, and she loves you. You got it, you got the job.	*joy*
4	Rachel	Oh, I did!	*surprise*
5	Mark	**Yes.**	*joy*

表 1.3: 正の感情を示す発話 "Yes" の例

目し，有向非巡回グラフニューラルネットワークを利用する手法を提案した [Shen et al., 2021]．

会話の履歴

対話の感情認識では，同じ発話であっても先行文脈に応じて異なる感情を示すことがあり，一連の会話の履歴が重要とされている [Jiao et al., 2019, Li et al., 2020a]．特に，"Yes" や "No"，"OK" などの短い発話は，対話の文脈に応じて様々な感情を示すことが知られている [Poria et al., 2019]．表 1.2 と表 1.3 を用いて，会話の履歴の重要性を示す．表 1.2 の 7 番目の発話 "Yes" は，それまでの会話の履歴から負の感情を示すが，表 1.3 の 5 番目の発話 "Yes" は正の感情を示す．このように，同じ発話であっても会話の履歴に応じて異なる感情を示すため，一連の会話の履歴は，対話の感情を識別する上で重要である．

これまでに多くの研究が会話の履歴を利用する手法を提案した [Jiao et al., 2019, Yang et al., 2019a, Li et al., 2020a,b, Liu et al., 2023]．Yang らは，複数の先行発話と識別対象の発話を連結し，事前学習済みの言語モデルに入力する方法を提案

図 1.1: 話者固有の特徴を示す例

した [Yang et al., 2019a]．入力系列の全体に注意を向けることができるため，会話全体の流れすなわち会話の履歴を利用することができる．

会話の履歴の利用を難しくする要因に，先行文脈の長さがある [Jiao et al., 2019, Li et al., 2020a,b]．識別対象の発話から先行発話への距離が遠い場合，遠い発話の情報を効果的に利用することは難しい．この課題に取り組む関連研究に，発話の内容を把握するネットワークと文脈を把握するネットワークを分けて階層的に組み合わせる手法がある [Jiao et al., 2019, Li et al., 2020a,b]．さらに，Shen らは，対話全体，近傍の発話，話者の依存関係に応じた注意機構を用意することで，効果的に周辺の発話に注意を向ける手法を提案した [Shen et al., 2020]．

話者固有の特徴

続いて，話者固有の特徴を示す．発言に込められる感情は話者の個性に依存するため, 同じ発話であっても話者に応じて異なる感情を示すことが知られている [Poria et al., 2019]．例えば，皮肉な表現を良く使う話者もいれば，そうでない話者も存在する．図 1.1 に示す，話者 A:“The order has been cancelled” に対する話者 B:“This is great” の返答は，仮に話者 B が頻繁に皮肉な発言を発する場合，返答を通して否定的な感情を示すと推測できる．一方で，皮肉な発言をあまり行わない場合，キャンセルした事実が話者 B にとって有益であり，肯定的な感情を表すと推測できる．

対話の感情認識におけるベンチマークデータは，話者の特徴を示す情報が不足する傾向にある [Poria et al., 2019]．そのため，これらの情報を利用する方法は認識性能の向上につながることが期待される．従来研究として，Li らは対話に登場する話者ごとに埋め込み表現を作成し，言語モデルに入力する方法を提案した [Li et al., 2020b]．また，Li らは話者固有の特徴を把握するために，同一会話内の複数の発話が同じ話者によるものかどうかを識別する手法を提案した [Li et al., 2020a]．最近の研究として Lee らは，事前に収集した話者の過去の発話を言語モデルに学習させることで，話者の発言パターンを記憶させる手法を提案した [Lee and Lee, 2022]．

研究課題	提案手法の位置付け
感情の種類の選択 感情ラベルの付与	Ekman の 6 感情を基にしたベンチマークデータセットを利用
識別モデルの構築	発話間の関係と会話の履歴に着目

表 1.4: 対話の感情認識の研究課題に対する提案手法の位置付け

常識的知識

常識的知識は，会話を理解し適切な応答を発するために必要な情報とされている [Zhou et al., 2018]．特に話し手聞き手の反応や意図に関する常識的知識は，話者の感情の動きを予測するのに役立つ．そこで従来研究として，Zhong らは発話のキーワードに関連する常識的知識を外部のデータベースから検索し，会話の履歴を利用するモデルに組み合わせる手法を提案した [Zhong et al., 2019]．また，Ghosal らは心理的状態や対話の状況に関連する常識的知識を利用する手法を提案した [Ghosal et al., 2020]．

1.3　提案手法の位置付け

本論文は，1.2 節に示す研究課題の中でも，特に識別モデルの構築に着目し，発話間の関係と会話の履歴を利用する新たな識別モデルを提案する．

まず，表 1.4 を用いて，1.2 節に示す対話の感情認識の課題に対する本研究の位置付けを示す．感情の種類の選択 (1.2.1 項) と感情ラベルの付与 (1.2.2 項) に対して，本論文は，識別モデルの構築に取り組む関連研究と同様に，1.2.1 項に示す Ekman の 6 感情を基にしたベンチマークデータセットを利用する．識別モデル構築 (1.2.3 項) では，研究課題の中でも発話間の関係と会話の履歴に着目する．発話間の関係や会話の履歴は，課題の解決によって高い認識性能を示すことが期待され，多くの従来研究の注目を集めている．そこで，本研究も従来研究と同様に，発話間の関係と会話の履歴に着目する．

次に，図 1.2 を用いて，発話間の関係に注目する提案手法と，会話の履歴に注目する提案手法の関連性および従来研究との位置付けを示す．対話の感情認識では，1.2.3 項に示す自分自身の感情の推移を表す自己依存と他者の発話が影響を与える他者依存の関係が感情に影響を与えるため，2 つの依存関係を利用する手法が多く提案された [Ghosal et al., 2019, Sheng et al., 2020, Shen et al., 2021]．これらの手法は，グラフニューラルネットワーク (GNN: Graph Neutral Networks) を用いて個々の発話の依存関係を利用することで高い認識性能を示す．

しかし，発話間の関係を利用する GNN は，発話の距離の利用が容易でない．表 1.1 の対話を用いて，発話の距離の重要性を示す．話者 B は 4 番目の発話で感情が変化する．これは直前の 3 番目の発話に同情したことが原因と考えられる．こ

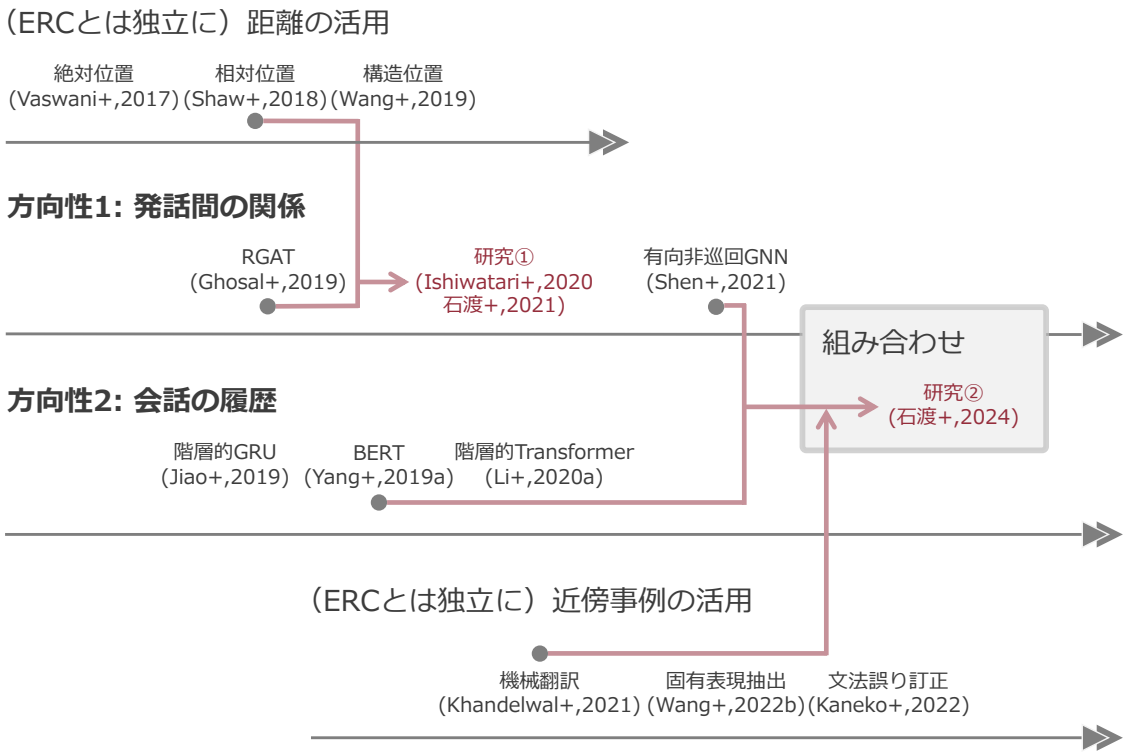

図 1.2: 発話間の関係に注目する提案手法と，会話の履歴に注目する提案手法の関連性および従来研究との位置付け

のように，話者の感情はしばしば発話から発話への距離に依存する．従って，発話の距離を利用することで，認識性能の向上が期待できる．

本論文は発話間の関係に加えて，対象の発話から周辺の発話への距離も利用する手法を提案する [Ishiwatari et al., 2020, 石渡 et al., 2021]．距離の影響を利用する一般的な方法として，絶対位置 [Vaswani et al., 2017] や相対位置 [Shaw et al., 2018] を基にした距離の情報を言語モデルに加える方法がある．発話から発話への依存関係を利用する GNN は，発話から発話への距離に基づく相対位置の適用が望ましい．そこで本論文は，相対位置に基づく距離の情報を，依存関係を利用する GNN に加える方法を提案する．提案手法を用いることで，自己依存と他者依存を含む発話間の関係と，発話の距離の両方を利用することができる．対話の感情認識のベンチマークデータセットによる評価実験を通して，距離を利用する提案手法の有効性を確認した．

次に, 会話の履歴に着目する提案手法を示す．距離を利用する提案手法 [Ishiwatari et al., 2020, 石渡 et al., 2021] や，最近の研究である DAG-ERC [Shen et al., 2021] は，GNN を用いて発話と発話の対の関係を利用し，高い認識性能を示した．しかし，これらの発話間の関係を利用する手法は，会話全体の一連の流れを考慮できない課題が存在する．1.2.3 項に示すように，対話の感情認識では，同じ発話であっても先行文脈に応じて異なる感情を示すことがあるため，一連の会話の履歴が重

要である [Jiao et al., 2019, Li et al., 2020b]．一般的な解決方法として，発話と発話を連結し対話全体を，トランスフォーマー (Transformer) [Vaswani et al., 2017] などの言語モデルに入力する方法がある [Yang et al., 2019a]．識別対象の発話だけでなく先行文脈にも注意を向けることができるため，会話の履歴を利用することが可能である．しかしながら，これらの会話の履歴を利用する手法は，入力系列の全体に注意を向けるため，逆に個々の発話の依存関係の利用が容易でない．1.2.3 項に示すように，話者の感情は発話間の関係や会話の履歴に依存するため，どちらの影響も重要である．

そこで，本論文は発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組みわせるアンサンブル手法を提案する．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いもの（近傍事例）を検索し，動的な重み付き線形和によって補強する事例ベース手法を提案する [石渡 et al., 2024]．提案手法を用いることで，発話間の関係と会話の履歴の両方の特徴を利用することができる．ベンチマークデータセットによる評価実験を通して，提案する事例ベース手法の有効性を確認した．

本論文は，対話の感情認識における発話間の関係と会話の履歴の課題に着目し，距離を考慮した発話間の関係を利用する手法を提案し，発話間の関係を利用するモデルに会話の履歴の観点で意味的に近い事例を組み合わせる手法を提案する．

1.4　論文の構成

本論文は，6 つの章で構成される．第 1 章 (本章) では，対話の感情認識の概要と課題，提案手法の位置付けを示した．第 2 章では，提案手法に関連する従来研究を示す．第 3 章では，対話の感情認識の問題設定とベンチマークデータ，評価方法を示す．第 4 章では，発話の距離を考慮した発話間の関係を利用する提案手法を説明する．第 5 章では，会話の履歴と発話間の関係を組み合わせる提案手法を示す．第 4 章と第 5 章に示す提案手法は，同じ問題設定かつ同じベンチマークデータを利用し評価する．そのため，問題設定とデータセット，評価方法の説明を第 3 章にまとめる．最後に，6 章で本論文の結論を示す．

第2章　関連研究

本論文は，1.3 節に示すように，対話の感情認識における発話間の関係と会話の履歴の課題に着目し，距離を考慮した発話間の関係を利用する手法を提案し，発話間の関係を利用するモデルに会話の履歴の観点で意味的に近い事例を組み合わせる手法を提案する．本章では，2 つの提案手法に関連する従来研究について議論する．はじめに，距離を考慮した発話間の関係を利用する提案手法の関連研究を示す．次に，会話の履歴と発話間の関係を組み合わせる提案手法の関連研究を示す．さらに，発話間の関係を利用するモデルと，会話の履歴を利用するモデルの特徴を示し，その違いを説明する．

2.1　発話の距離を考慮した発話間の関係

本論文は，対象の発話から周辺の発話への距離を，発話間の関係を利用する GNN に加える方法を提案する．はじめに GNN に関する研究を説明する．次に，自己依存と他者依存を含む発話間の関係を利用する従来手法を示し，距離の情報を活用する従来手法を示す．

2.1.1　グラフニューラルネットワーク (GNN)

GNN は，自然言語処理の分野だけでなく画像処理など様々な問題に応用され，著しい注目を集めている．いくつか種類のある GNN の中でも，ノードとノードの結合関係を表す隣接行列を利用したグラフ畳み込みネットワーク (GCN: Graph Convolutional Networks) [Kipf and Welling, 2016] をはじめ，関係グラフ畳み込みネットワーク (RGCN: Relational Graph Convolutional Networks) [Schlichtkrull et al., 2018] やグラフアテンションネットワーク (GAT: Graph Attention neTworks) [Veličković et al., 2017] が盛んに利用されている．提案手法は，発話間の関係を利用するために，RGCN と GAT を組み合わせた RGAT を用いる．RGCN はノード間の関係の種類ごとに GCN を用意するため，自己依存と他者依存など依存関係の種類ごとにネットワークを構築することができる．また GAT を用いることで，ノード間の類似度を計算し，関連性のあるノードに注意を向けることができる．

2.1.2　発話間の関係を利用する手法

次に，発話間の関係を利用する従来手法を示す．対話の感情認識では，1.2.3 項に示すように，自分自身の感情の推移を表す自己依存と他者の発話が影響を与える他者依存の関係が感情に影響を与えるため，2 つの依存関係を利用する手法が多く提案された [Ghosal et al., 2019, Sheng et al., 2020, Shen et al., 2021]．これらの手法は，GNN を用いて個々の発話の依存関係を利用することで高い認識性能を示した．従来研究の中で，Sheng らの手法 [Sheng et al., 2020] は，自己依存と他者依存を考慮する GAT を構築し高い認識性能を示した．

提案手法に最も関連のある手法として，DialogueGCN [Ghosal et al., 2019] がある．DialogueGCN は，自己依存と他者依存の関係を利用するために RGAT を活用し，当時の高い認識性能を示した．Ghosal らは，対話に登場する話者ごとに自己依存と他者依存の影響の度合いが異なると考え，話者ごとに自己依存と他者依存を区別する RGAT を構築した．しかし，話者の数が増えると，依存関係の種類が増えパラメータの数も増加する．そこで提案手法は，パラメータ数の削減を目的として，話者ごとに自己依存と他者依存を区別しない RGAT を構築する．加えて，DialogueGCN を含む GNN ベースのモデルは，発話の距離の利用が容易でない．本論文は，発話の距離の利用を目的として，依存関係の種類に応じた位置の埋め込みを新たに作成し，RGAT に加える手法を提案する．

2.1.3　距離を活用する手法

本論文は, 対象の発話から周辺の発話への距離を, 発話間の関係を利用する RGAT に加える方法を提案する．関連研究に, 位置埋め込みを Transformer [Vaswani et al., 2017] に加える方法がある．Transformer は注意機構を基に構成されるため，時系列情報の利用が容易ではない．そこで，絶対位置 [Vaswani et al., 2017] や相対位置 [Shaw et al., 2018]，構造位置 [Wang et al., 2019] に基づく位置埋め込み手法が提案されている．RGAT を含む GNN も同様に時系列情報の利用が容易ではない．そのため Ingraham らはタンパク質の設計に際し，タンパク質の相対的な位置を GNN のエッジに付加するモデルを提案した [Ingraham et al., 2019]．

2.2　会話の履歴と発話間の関係の組み合わせ

本論文は会話の履歴を利用するモデルと，発話間の関係を利用するモデルを組みわせるアンサンブル手法を提案する．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いもの（近傍事例）を検索し，動的な重み付き線形和によって補強する事例ベース手法を提案する．本節は，はじめに，会話の履歴を利用する従来研究を示す．続いて，発話間の関係を利用する従来研究を示す．なお，本節で示す発話間の関係を利用する従来手法は，提案手法のアンサンブルに利用

する手法であり，2.1.2 項に示す従来研究よりも最新の研究である．次に，近傍事例を活用する従来手法を説明し，提案手法との関連性を示す．

2.2.1 会話の履歴を利用する手法

はじめに，会話の履歴を利用する従来研究を示す．対話の感情認識では，1.2.3 項に示すように，同じ発話であっても会話の履歴に応じて異なる感情を示すため，会話の履歴を利用する手法が多く提案された [Poria et al., 2017, Jiao et al., 2019, Majumder et al., 2019, Yang et al., 2019a, Li et al., 2020a]．これらの手法は，リカレントニューラルネットワーク (RNN: Recurrent Neural Networks) や Transformer を用いて，複数発話の系列情報を利用することで高い認識性能を示した．会話の履歴を利用する代表的な方法に，Poria らの手法がある [Poria et al., 2017]．Poria らの手法は，畳み込みニューラルネットワーク (CNN: Convolutional Neural Networks) [Karpathy et al., 2014] を用いて発話の内容を示す特徴量ベクトルを作成し，双方向 Long Short Term Memory(LSTM) [Hochreiter and Schmidhuber, 1997] を用いて発話間の関連性を取得し，会話の文脈を把握する手法である．また Jiao らの手法は，単一発話の内容を把握する双方向ゲート付リカレントニューラルネットワーク (GRU: Gated Recurrent Unit) [Chung et al., 2014] と，会話の文脈を把握する双方向 GRU を用意し，階層的に組み合わせる手法である [Jiao et al., 2019]．DialogueRNN [Majumder et al., 2019] は，GRU と注意機構を用いて関連のある発話に焦点を当て，対話の感情認識の性能向上に大きく貢献した．

提案手法に最も関連のある手法として，Yang らの手法 [Yang et al., 2019a] がある．この手法は，複数の発話の間に [SEP] トークンを挿入して連結し，対話全体を事前学習済み BERT (Bidirectional Encoder Representations from Transformers) [Devlin et al., 2018] に入力する．各発話の各トークンの特徴量を取得し，最大値プーリングを用いて識別対象の発話の特徴量を取得する．取得した対象発話の特徴量をフィードフォワードネットワーク (FFN: Feed Forward Networks) に入力して感情ラベルの確率値を算出する．複数の発話を同時に入力することで，発話の系列すなわち会話の履歴を利用することができる．提案手法は，Yang らの手法 [Yang et al., 2019a] を参考に，識別対象の発話と先行文脈を連結し入力する RoBERTa [Liu et al., 2019b] を，会話の履歴を利用するモデルとして利用する．

2.2.2 発話間の関係を利用する手法

提案手法は，DAG-ERC [Shen et al., 2021] または DialogueCRN [Hu et al., 2021] を，発話間の関係を利用する手法として用いる．DAG-ERC は，話者自身の離れた発話からの影響（自己依存）と，他者の近い発話からの影響（他者依存）を考慮するために，GAT を拡張した有向非巡回グラフニューラルネットワークを新たに

構築し，対話の感情認識の性能向上に大きく貢献した [Shen et al., 2021]．また，DialogueCRN は，対話の状況に応じた相互作用と自身の発言による自己依存を利用し，高い認識性能を示した [Hu et al., 2021]．提案手法は，高い認識性能を示す DAG-ERC [Shen et al., 2021] または DiaglogueCRN [Hu et al., 2021] を，近傍事例を組み合わせる対象の発話間の関係を利用するモデルとして利用する．

2.2.3　近傍事例を活用する手法

本論文は，発話間の関係を利用するモデルに，会話の履歴の観点で意味的に近いものを過去の会話から検索し組み合わせる事例ベース手法を提案する．提案手法に最も関連のある事例ベース手法として，機械翻訳タスクに近傍事例を活用する手法 [Khandelwal et al., 2021] がある．この手法はまず，訓練データ (対訳文) の原言語文を事前学習済みニューラル機械翻訳 (NMT: Neural Machine Translation) モデルに入力し，翻訳の各時刻すなわち各単語の位置における中間表現を取得する．NMT モデルのデコーダの最終層から得られる特徴量ベクトルを中間表現として用いる．続いて推論時に，評価データの原言語文を事前学習済み NMT モデルに入力し，翻訳中の時刻 τ の単語の位置における中間表現を取得する．訓練データと同様に，NMT モデルのデコーダから得られる特徴量を中間表現として用いる．取得した時刻 τ の単語の位置における中間表現と，訓練データから事前に計算した各中間表現との距離を計算し，距離の近い事例 (近傍事例) を検索する．検索した近傍事例の距離と，事例が示す単語の出現頻度に基づいた単語予測分布を作成し，時刻 τ における NMT モデルの単語予測分布と重み付き線形和によって組み合わせる．追加の学習をせずに，近傍事例を従来の NMT モデルに組み合わせることで，翻訳性能が大幅に改善することが報告されている [Khandelwal et al., 2021]．

また, 近傍事例を活用する手法は, 機械翻訳 [Khandelwal et al., 2021, Zheng et al., 2021, Jiang et al., 2021, Wang et al., 2022a] だけでなく, 固有表現抽出 [Wang et al., 2022b]，文法誤り訂正 [Kaneko et al., 2022] などの幅広い問題設定に応用され，有効性が示されている．しかし，対話の感情認識に，近傍事例を応用する手法は存在しない．そこで本研究は，対話の感情認識タスクに近傍事例を初めて応用する．

近傍事例の距離や教師ラベルを利用するだけでなく，近傍事例の特徴量を作成し活用する手法も注目を集めている．He らの手法 [He et al., 2021] は，機械翻訳タスクにおいて, 近傍事例に付与された参照文から単語埋め込みと位置埋め込みを用いて特徴量を作成し，NMT モデルのデコーダに追加で入力する．また Borgeaud らの手法 [Borgeaud et al., 2022] は，入力テキストをある程度の長さ (チャンク) で分割し，分割したチャンクごとに近傍事例を検索する．検索した近傍事例の文章から Transformer を用いてトークンごとの特徴量を作成し，交差注意 (CA: Cross Attention) を用いて言語モデルに反映する．これらの従来手法は，近傍事例の特徴量を利用する点で提案手法と関連するが，注意機構を用いてモデルに反映させる点で提案手法と異なる．提案手法は，近傍事例の特徴量を入力し，発話間の関

係を利用するモデルの確率分布と近傍事例の確率分布を組み合わせるための重み係数を導出する．

2.3 発話間の関係を利用するモデルと会話の履歴を利用するモデルの特徴

本論文は，4 章に示す提案手法において，RGAT [Schlichtkrull et al., 2018, Veličković et al., 2017] を発話間の関係を利用するモデルとして利用し，5 章に示す提案手法において，Transformer [Vaswani et al., 2017] を基にした RoBERTa [Liu et al., 2019b] を会話の履歴を利用するモデルとして利用する．本節は，RGAT と Transformer によるモデリングの特徴を示し，その違いを説明する．図 2.1 は，先行発話の数が 4 で，10 番目の発話を識別対象としたときの，RGAT と Transformer の各層の特徴量ベクトルを作成する方法を表す．白丸は話者 A の発話の特徴量を，黒丸は話者 B の発話の特徴量を示す．白四角は話者 A のトークンの特徴量を，黒四角は話者 B のトークンの特徴量を示す．表 2.1 は，注意機構と特徴量の作成手順，研究課題への適用に関する RGAT と Transformer の特徴を示す．

2.3.1 発話間の関係を利用する RGAT

はじめに，RGAT の基となる GAT [Veličković et al., 2017] のモデリング方法を示す．図 2.1 に示すように，GAT は識別対象の発話の特徴量ベクトルを作成する際，隣接発話とグラフのエッジを構築し，その特徴量ベクトルを利用する．具体的には，10 番目の 2 層目の発話の特徴量ベクトルを作成する際，隣接する 4 つの先行発話の 1 層目の特徴量ベクトルを利用する．さらに，隣接発話の 1 つである 6 番目の発話の 1 層目の特徴量ベクトルを作成する際，6 番目の発話に隣接する 1 つ前の層の発話の特徴量ベクトルを利用する．GAT は注意機構を用いることで，隣接発話の中で関連性のある発話に注意を向ける．また，1 番目の 1 層目の発話から識別対象の最終層の発話まで，順に特徴量ベクトルを作成する．層数を重ねることで，層数分のエッジによって結ばれた隣接発話の特徴を反映することができる．

次に，RGAT のモデリング方法を示す．図 2.1 に示すように，RGAT は隣接発話と関係の種類ごとにグラフのエッジを構築し，その特徴量を利用する．従来研究として Ghosal らは，同じ話者による発話からの影響を表す自己依存と，違う話者による発話からの影響を表す他者依存の関係に基づき，エッジを構築した [Ghosal et al., 2019]．RGAT は 2 層目の話者 A による識別対象発話の特徴量を作成する際，自身による発話か他者による発話かに応じて，1 層目の 4 つの先行発話[1]を区別し，その特徴量を利用する．RGAT は，GAT と同様に，各層の各発話の特徴量

[1]本稿はエッジの種類ごとにハイパーパラメータである窓幅を設定する．詳細は 4.2.2 項に示す．

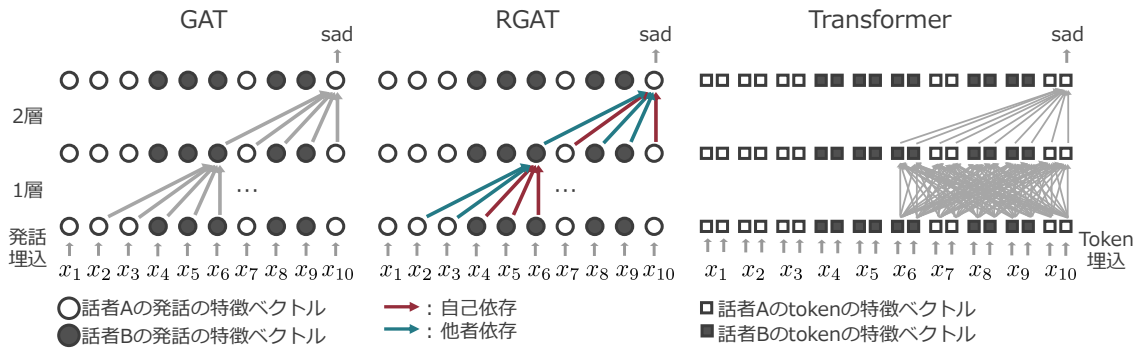

図 2.1: 発話間の関係を利用する RGAT と，会話の履歴を利用する Transformer の比較.

特徴	GAT	RGAT	Transformer
注意機構	局所的なノードに注意	エッジの種類ごとの局所的なノードに注意	入力系列の全ノードに注意
特徴量の作成手順	各ノードを順に更新	各ノードを順に更新	全ノードを同時に更新
研究課題への適用	-	依存関係の種類に基づく発話間の関係に適用	トークン系列の関連性を利用する会話の履歴に適用

表 2.1: 注意機構と特徴量の作成手順，研究課題への適用に関する，発話間の関係を利用する RGAT と会話の履歴を利用する Transformer の特徴の比較.

ベクトルを順に作成する．以上のように，RGAT はエッジの種類を考慮して各発話の特徴量を作成するため，自己依存と他者依存が感情に影響を与える発話間の関係のモデリングに適している．本論文は，発話間の関係を利用するモデルとして，RGAT を活用する．

2.3.2　会話の履歴を利用する Transformer

次に, Transformer のモデリング方法を示す．Transformer は入力したトークン系列の全てのトークンに注意を向け，その関連性を利用する．従来研究として Yang らは，識別対象の発話とその先行発話を連結し BERT [Devlin et al., 2018] に入力することで，その関連性を利用した [Yang et al., 2019a]．図 2.1 に示すように，Transformer は，識別対象発話の特徴量を作成する際，連結した 4 つの先行発話と対象発話のトークン系列の関連性を利用する．Transformer は，GAT [Veličković et al., 2017] と異なり，入力系列の全ての特徴量を各層で同時に作成する．以上のように，Transformer は入力した系列発話の全てのトークンの関連性を利用するため，会話の履歴のモデリングに適している．本論文は，会話の履歴を利用するモデルとして，Transformer を基にした RoBERTa を活用する．

第3章　対話における感情認識

3.1　問題設定

はじめに対話の感情認識の問題設定を示す．この問題設定は，対話における各発話 $(x_1, x_2, \cdots, x_N)$ の感情ラベル $(y_1, y_2, \cdots, y_N)$，$y_n \in \mathbb{Y}$ を認識する．N は1つの対話に現れる発話の数を示す．$\mathbb{Y}$ は感情ラベルの集合を示し，感情ラベルのラベル数を C とする．また，1つの対話に登場する話者を s_v $(v = 1, \cdots, V)$ とする．V は話者の人数を示す．図 3.1 に，発話の数が N の対話における n 番目の発話の感情ラベルを識別する例を示す．対話の感情認識では，全ての対話の全ての発話の感情ラベルを識別する．

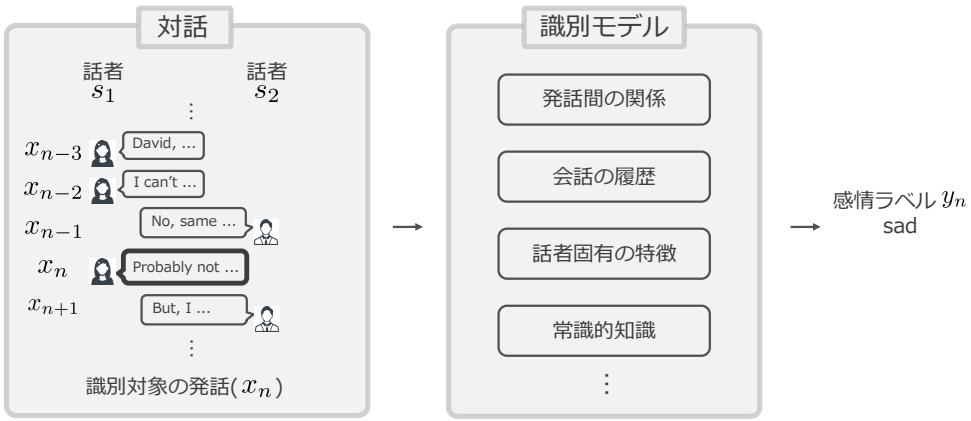

図 3.1: 対話の感情認識の概要．n 番目の発話の感情を識別する例．識別モデルに対話を入力し，感情ラベルを識別する．

データセット	対話数			発話数			ラベル数
	訓練	検証	評価	訓練	検証	評価	
IEMOCAP	100	20	31	4860	950	1623	6
MELD	1038	114	280	9989	1109	2610	7
EmoryNLP	713	99	85	9934	1344	1328	7

表 3.1: IEMOCAP，MELD，EmoryNLP ベンチマークデータセットの割合．

3.2　ベンチマークデータセット

本論文は，対話の感情認識における 3 つのベンチマークセットを用いる．それぞれのデータセットの訓練セット，検証セット，評価セットの対話数と発話数とラベル数を表 3.1 に示す．

IEMOCAP [Busso et al., 2008]　は 2 人の話者が，1 対 1 の会話を行う様子を収録した映像と音声の書き起こしからなるデータセットである．話者は，適切な感情が自然に引き出されるような仮想的なシナリオに沿って会話を行うため，会話の台本や各発話の感情が制御されたデータセットである．この場合，台本に記された感情を感情ラベルとして利用することが可能である．しかし，実際の会話は想定したシナリオと異なる展開を示すことがあり，シナリオの感情を適切に反映しない．そこで，シナリオの感情を利用する方法に代わり，書き起こした会話に事後的に評価者が感情ラベルを付与する方法が採用されている．3 人の異なる評価者は，収録した映像と書き起こした発話を基に感情ラベルを評価する．評価者間のラベルの一致度を計算するために，Fleiss' Kappa [Fleiss, 1971] を利用する．IEMOCAP の Fleiss' Kappa は 0.4 である．各発話には，*happy*, *sad*, *neutral*, *angry*, *excited*, *frustrated* の感情ラベルのうち 1 つが付与されている．IEMOCAP の訓練セット，検証セット，評価セットにおける，各感情ラベルの出現頻度を表 3.2 に示す．

IEMOCAP	ラベルの出現頻度		
	訓練セット	検証セット	評価セット
happy	425	79	144
sad	671	168	245
neutral	1084	240	384
angry	789	144	170
excited	664	78	299
frustrated	1227	241	381

表 3.2: IEMOCAP における各感情ラベルの出現頻度.

MELD [Poria et al., 2018]　は，複数の俳優が登場する Friends という TV ドラマの，一部シーンを切り取った映像と音声の書き起こしからなるデータセットである．MELD は，Emotionlines [Chen et al., 2018] を拡張したデータセットで，1 つの対話に複数の話者が登場する．約 42% の発話は 5 単語以下の短い発話から構成される．異なる 3 人の評価者によってラベルを付与し，多数決でラベルを決定する．3 人の評価者間で付与したラベルが全て異なる場合，その発話をデータセットから除く．MELD の Fleiss' Kappa [Fleiss, 1971] は 0.43 である．各発話に

は，*neutral*, *joy*, *sadness*, *anger*, *surprise*, *fear*, *disgust* の感情ラベルのうち 1 つが付与されている．MELD の訓練セット，検証セット，評価セットにおける，各感情ラベルの出現頻度を表 3.3 に示す．

MELD	ラベルの出現頻度		
	訓練セット	検証セット	評価セット
neutral	4710	470	1256
joy	1743	163	402
sadness	683	111	208
anger	1109	153	345
surprise	1205	150	281
fear	268	40	50
disgust	271	22	68

表 3.3: MELD における各感情ラベルの出現頻度．

EmoryNLP [Zahiri and Choi, 2018] は TV ドラマ Friends から，一部のシーンを切り取り収集したデータセットである．MELD とデータサイズとラベルの種類が異なり，*neutral*, *sad*, *mad*, *scared*, *powerful*, *peaceful*, *joyful* のうち 1 つが付与される．また IEMOCAP と MELD と異なり，異なる 4 人の評価者によってラベルを付与し，多数決で決定する．評価者間のラベルの一致度 Fleiss' Kappa [Fleiss, 1971] は 0.14 である．EmoryNLP の訓練セット，検証セット，評価セットにおける，各感情ラベルの出現頻度を表 3.4 に示す．neutral, positive (*powerful*, *peaceful*, *joyful*)，negative (*sad*, *mad*, *scared*) の 3 感情の出現頻度は約 30%, 約 40%, 約 30% の割合を占める．従って，neutral, positive, negative の割合がバランスされたデータセットである．

EmoryNLP	ラベルの出現頻度		
	訓練セット	検証セット	評価セット
neutral	3034	393	349
sad	671	75	98
mad	1076	143	113
scared	1285	178	182
powerful	784	134	145
peaceful	900	132	159
joyful	2184	289	282

表 3.4: EmoryNLP における各感情ラベルの出現頻度．

3.3　評価方法

対話の感情認識の評価には，重み付き F1 値が用いられる [Ghosal et al., 2019, Shen et al., 2021]．重み付き F1 値は，分類問題において，教師ラベルと予測ラベルとの一致を測る自動評価尺度である．F1 値は適合率と再現率の調和平均によって計算され，$[0 \leq \mathrm{F1} \leq 1]$ の範囲の値を示し，1 に近づくほど高い一致率を示す．各ラベルにおける F1 値は，下式で計算する．

$$\mathrm{F1} = \frac{2 * \text{適合率} * \text{再現率}}{\text{適合率} + \text{再現率}} \tag{3.1}$$

重み付き F1 値は，式 (3.1) に示す F1 値をラベルごとに計算し，各ラベルの出現回数で重み付けした平均を計算する．

対話の感情認識のベンチマークデータセットは，表 3.2，表 3.3，表 3.4 に示すように各感情ラベルの出現回数に偏りがある．多くの従来研究は，出現回数の偏りによる影響を削減するために，各ラベルの出現回数によって重み付けされた F1 値（重み付き F1 値）を評価指標として利用する [Ghosal et al., 2019, Shen et al., 2021].

第4章　発話の距離を考慮した発話間の関係

4.1　研究の概要

本章は，発話の距離を考慮した発話間の関係を利用する識別モデルを提案する．対話の感情認識では，対話における各発話の内容に加えて，発話間の関係が話者の感情に大きな影響を与えることが知られている [Poria et al., 2019]．発話間の関係の中でも，自身の発話からの影響 (自己依存) と他者の発話からの影響 (他者依存) が重要である [Ghosal et al., 2019, Shen et al., 2021]．表 1.1 を用いて，2 つの依存関係の重要性を示す．話者 A は長い間就職先が見つからないため，一連の発話で常に負の感情を抱いている．これは自己依存の例を示し，自分自身の感情の推移を表す．一方で，話者 B の 4 番目の感情は，直前の話者 A の状況に同情し，負の感情を抱いている．これは他者依存の例を示し，他者の発話が自身の感情に影響を与える性質を持つ．

Ghosal らは自己依存と他者依存の関係を利用するために，RGAT[1]を用いて，当時の世界最高峰の認識性能を示す手法を提案した [Ghosal et al., 2019]．この手法は，ノードに各発話の特徴量を，エッジに発話間の関係を，エッジの種類に依存関係の種類を設定し，有向グラフを構築する．しかしながら，RGAT を含む GNN は会話中の発話の距離を利用できない課題がある．表 1.1 を用いて，発話の距離の重要性を示す．話者 B は 4 番目の発話で，感情が変化する．これは，1 番目や 2 番目の発話ではなく，直前の 3 番目の発話に同情したことが原因と考えられる．従って，発話から発話への距離の影響を利用することで，対話の感情認識の認識性能の向上が期待できる．

距離の影響を利用する一般的な方法として，発話の絶対位置 [Vaswani et al., 2017] や相対位置 [Shaw et al., 2018] を基にした距離の情報を，GNN に加える方法がある．絶対位置は GNN のノード (発話) に，相対位置はエッジ (発話間の関係) に加えられる．一方で提案手法は，Ghosal らの手法 [Ghosal et al., 2019] を参考に，自己依存と他者依存の利用を目的として，依存関係の種類に応じた RGAT を用いる．従って，絶対位置や相対位置ではなく，依存関係の種類に応じた距離の情報を用いることで，認識性能の向上が期待できる．

[1]RGCN: Relational Graph Convolutional Networks [Schlichtkrull et al., 2018] と GAT: Graph Attention neTworks [Veličković et al., 2017] を組み合わせたモデル

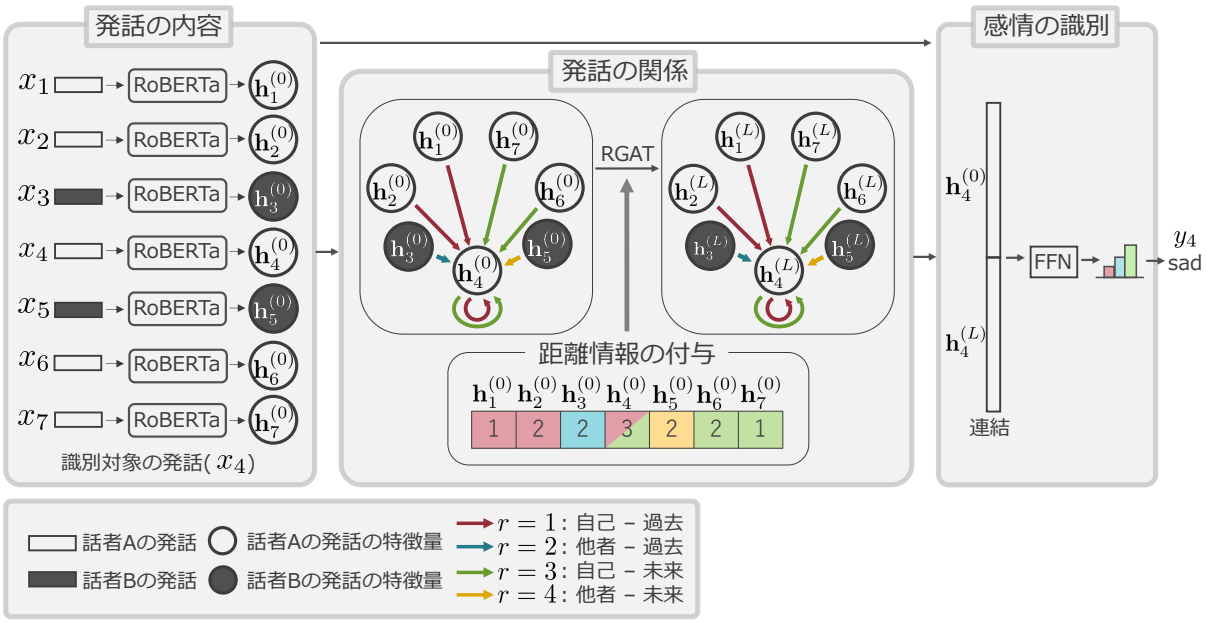

図 4.1: 提案手法の全体図．はじめに RoBERTa を用いて発話の内容を示す特徴量を作成する．次に，RGAT を用いて発話間の関係を考慮した特徴量を作成する．$\mathbf{h}_i^{(l)}$ は発話 x_i の l 層目の特徴量を表す．次に，RGAT に発話の距離の情報を追加する．最後に発話の特徴量と発話間の関係を考慮した特徴量を連結し，FFN を用いて感情ラベルを識別する．図は発話 x_4 の感情ラベルを識別する例を示す．

本論文は，依存関係の種類に応じた位置の埋め込みを新たに作成し，RGAT に加える方法を提案する．提案手法を用いることで，自己依存と他者依存を含む発話間の関係と，発話の距離の両方を利用できる．評価実験において，ERC における 3 つのベンチマークデータセットのうち，2 つのデータセットで従来手法を上回り，高い認識精度を示した．さらに，依存関係の種類に応じた距離の情報が，対話の感情認識の精度向上に貢献することも確認した．

4.2　提案手法

提案手法は，発話の内容，距離を考慮した発話間の関係，感情ラベルの識別の 3 つで構成される．概要を図 4.1 に示す．本手法は，Ghosal らの手法 [Ghosal et al., 2019] を参考に，発話間の関係の利用を目的として RGAT を構築する．さらに，RGAT では取得が容易でない発話の距離の利用を目的として，距離の情報を加える手法を提案する．

4.2.1 発話の内容

まず，BERT [Devlin et al., 2018] を用いた発話の内容を示す特徴量ベクトルを作成する方法 [Luo and Wang, 2019] を参考に，発話のトークンごとに特徴量を作成する．各発話 $x_1, x_2, \cdots, x_N$ を RoBERTa の tokenizer を用いて，トークンごとに分割する．次に，事前学習モデル RoBERTa-large[2] [Liu et al., 2019b] に分割したトークンを入力し，発話の内容を考慮した特徴量を作成する．最後に，Max-pooling を通じて発話 x_i の特徴量 $\mathbf{h}_i^{(0)} \in \mathbb{R}^d$ を得る．d は発話の内容を示す特徴量の次元数を示す．事前学習済み RoBERTa は，学習ステップでファインチューニングを行う．

4.2.2 距離を考慮した発話間の関係

次に，4.2.1 項で作成した特徴量 $\mathbf{h}_i^{(0)}$ を RGAT に入力し，発話間の関係を考慮した特徴量を作成する．本手法は，自己依存と他者依存の関係を区別するために，RGAT を用いて関係の種類に応じたネットワークを用意する．また，注意機構を導入し，関連性のある発話に注意を向ける．さらに，発話の距離の利用を目的として，依存関係の種類に応じた距離の情報を加える方法を提案する．

グラフの構造

はじめにグラフの構造を定義する．発話の特徴量をグラフのノードとし，発話間の関係を 2 つのノード (発話の特徴量) 間を結ぶエッジとする．$r \in \mathcal{R}$ はエッジの種類 (依存関係の種類) を示す．

ノード グラフのノードは，各発話の特徴量で表現する．ノードは，発話の特徴量 $\mathbf{h}_i^{(0)}$ を初期値とする．複数の層の RGAT を重ねることで，隣接する発話の特徴量を複数回集計する．L 層重ねた発話 x_i の特徴量を $\mathbf{h}_i^{(L)}$ とする．

エッジの種類 発話間の関係ごとにグラフを構築する Ghosal らの手法 [Ghosal et al., 2019] を参考に，(a) 話者の関係と (b) 時間の関係に基づきエッジの種類を決定する．(a) 話者の関係 : 発話間の関係を自己依存と他者依存のいずれかに割り当てる．話者 s_v による発話 x_i に対して，同じ話者 s_v の発話 (x_i を含む) を自己依存とする．一方で，話者 s_v による発話 x_i に対して，違う話者 $s_{k\neq v}$ の発話を他者依存とする．(b) 時間の関係 : 発せられた時間によって，エッジの種類を割り当てる．すなわち発話 x_i に対して，発話 x_j が先に発せられたか (過去)，あるいは後に発せられたか (未来) に応じて種類を分ける．一般的に，リアルタイムの対話では，未来の発話を利用できないが，対話の感情認識はデータが全て揃ったオフラインを想定した応用も考えられるため，本研究は未来の発話も利用する．

[2] https://github.com/pytorch/fairseq

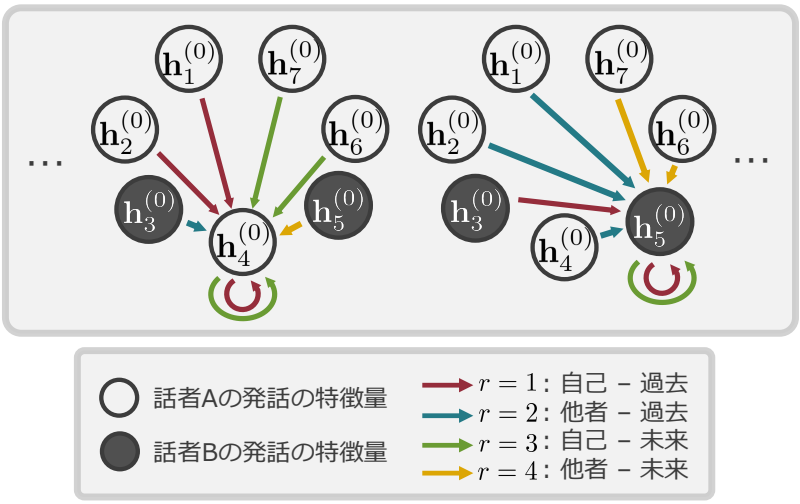

図 4.2: グラフの構築方法．対話における全ての発話に対して，エッジの種類に基づくグラフを構築する．左は発話 x_4 の特徴量 $\mathbf{h}_4$ を基準にしたとき，右は発話 x_5 の特徴量 $\mathbf{h}_5$ を基準したときのエッジの種類を示す．

以上より，(a) 話者の関係と (b) 時間の関係に基づき，(1) 自己 - 過去，(2) 他者 - 過去，(3) 自己 - 未来，そして (4) 他者 - 未来，の計 4 種類のエッジを設定する．エッジの種類を $\mathcal{R} = \{1, 2, 3, 4\}$ とする．

次に，エッジの種類について，Ghosal らの手法 [Ghosal et al., 2019] と本手法の違いを説明する．Ghosal らは自己依存と他者依存の影響の度合いが話者ごとに異なると考え，登場する話者の数に応じて，(a) 話者の関係と (b) 時間の関係に基づくエッジを用意した．例えば，ある対話に s_1, s_2 の話者 2 人が登場する場合，(1) 話者 s_1 - 自己 - 過去，(2) 話者 s_2 - 自己 - 過去，$\cdots$，(8) 話者 s_2 - 他者 - 未来の，計 2(話者の数) $\times$ 4(関係の種類) $= 8$ 種類のエッジを用意する．Ghosal らの手法は，話者ごとにネットワークを区別するため，特定の話者に関する影響を考慮することができる．しかしながら，対話に登場する話者の数が増えると，エッジの種類が増えパラメータの数も増加する．そこで本手法は，パラメータの数を削減するために，話者ごとに関係を区別せず，(a) 話者の関係と (b) 時間の関係に基づく 4 種類の関係を用いる．

グラフの構築　対話における全ての発話に対して，4 種類の関係に基づいたグラフを構築する．図 4.2 を用いて，グラフ構築の例を示す．図 4.2 は，4 番目の発話 (左) と 5 番目の発話 (右) を基準にしたときの，隣接する発話との関係を示す．例えば，4 番目の発話を基準にした 1 番目の発話は，過去の自身の発話なので，赤色で示す (1) 自己 - 過去の関係 ($r = 1$) を表す．また，6 番目の発話は未来の自身の発話なので，緑色で示す (3) 自己 - 未来の関係 ($r = 3$) を表す．次に，5 番目の発話を基準にした図 4.2 の右側の例を考える．5 番目の発話を基準にした 2 番目の発

話は，過去の違う話者の発話なので，青色で示す (2) 他者 - 過去の関係 ($r = 2$) を表す．このように，対話における全ての発話 x_i に対して，隣接する発話 x_j と，話者と時間の関係に基づく 4 種類の関係を構築する．

エッジの上限　エッジを結ぶ発話の上限 (窓幅) を定める．過去の窓幅を p，未来の窓幅を f としたとき，発話 x_i と p 個の過去の発話，f 個の未来の発話をエッジで結ぶ．窓幅を小さく設定する場合，小さな隣接グループしか注目することができない．一方で窓幅を大きくする場合，高い計算コストが必要となる．従って，適当な窓幅 p と f の設定が求められる．本手法は，ハイパーパラメータの数を削減するために，過去の窓幅 p と未来の窓幅 f を同じ値 $p = f$ に設定する．

隣接発話の集合　発話 x_i の隣接する発話の集合を定める．エッジの種類 r で発話 x_i に隣接する発話の集合を $\mathcal{N}_r(i)$ とする．例えば，図 4.2 において窓幅 $p = 3$ で 5 番目の発話 (右側) を基準にしたとき，青色で示すエッジの種類 $r = 2$ で発話 x_5 に隣接する発話の集合は，$\mathcal{N}_2(5) = \{1, 2, 4\}$ となる．

RGAT

RGCN [Schlichtkrull et al., 2018] と GAT [Veličković et al., 2017] に基づき，隣接する発話の特徴を加味した，$(l + 1)$ 層目の発話 x_i の特徴量 $\mathbf{h}_i^{(l+1)}$ を作成する．次に示す，発話 x_i と発話 x_j 間の注意機構を加味した重み係数 $\alpha_{ij}^{(l)}$ により，隣接する発話の重み付けを行う．エッジの種類 r で隣接する発話 $\mathcal{N}_r(i)$ の特徴を，全てのエッジの種類 $\mathcal{R}$ で集計した特徴量を下式で示す．

$$\mathbf{h}_i^{(l+1)} = \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r(i)} \alpha_{ij}^{(l)} W_r^{(l)} \mathbf{h}_j^{(l)} \tag{4.1}$$

$W_r^{(l)}$ は，エッジの種類 r ごとに用意した RGAT のパラメータである．RGAT の層数を L 回重ねることで，L 回分のエッジによって結ばれた隣接ノードの特徴を反映することができる．さらに本手法は式 (4.1) に Multi Head Attention を適用し，Layer Normalization も加える．

エッジの重み係数

式 (4.1) における，注意機構を用いたエッジの重み係数 $\alpha_{ij}^{(l)}$ を導入する．本手法は，GAT [Veličković et al., 2017] と同様に注意機構を利用する．発話 x_i と，エッジの種類 r で隣接する発話の集合 $\mathcal{N}_r(i)$ の発話 x_j の，エッジの重み係数 $\alpha_{ij}^{(l)}$ を下式で示す．

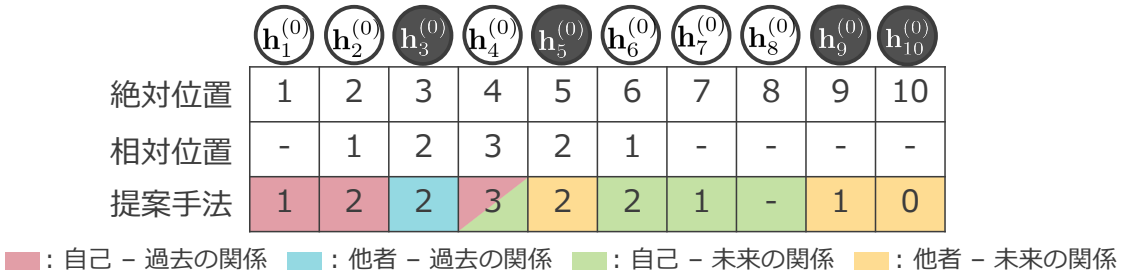

図 4.3: 提案する位置の埋め込みの例．提案手法は，エッジの種類ごとに用意した相対位置を用いる．背景色は x_4 を基準とした時のエッジの種類を示す．提案手法の値は，窓幅 $p = 3$ のときの発話 x_4 からの位置の埋め込みを示す．

$$\alpha_{ij}^{(l)} = \frac{\exp(\beta_{ij}^{(l)})}{\sum_{\bar{j}\in\mathcal{N}_r(i)} \exp(\beta_{i\bar{j}}^{(l)})} \tag{4.2}$$
$$\beta_{ij}^{(l)} = \mathrm{LeakyReLU}\big((\mathbf{a}_r^{(l)})^T [W_r^{(l)}\mathbf{h}_i^{(l)} || W_r^{(l)}\mathbf{h}_j^{(l)}]\big)$$

ただし，式 (4.1) に示すように，j はエッジの種類 r で発話 x_i に隣接する集合 $\mathcal{N}_r(i)$ の要素である．従って，$j \in \mathcal{N}_r(i)$ の関係から，(i, j) の組が定まるとエッジの種類 r も定まる．また，$W_r^{(l)}$ は注意機構のパラメータを，$\mathbf{a}_r^{(l)}$ は学習可能なベクトルを，$\cdot^T$ は転置を示す．LeakyReLU (Leaky Rectified Linear Unit) 活性化関数を通じて得た $\beta_{ij}^{(l)}$ を，集合 $\mathcal{N}_r(i)$ における edge softmax[3] [Veličković et al., 2017] で正規化し，重み係数 $\alpha_{ij}^{(l)}$ を得る．

提案手法の距離

RGAT では取得が容易でない発話の距離の利用を目的として，エッジの種類に応じた位置の埋め込みを新たに作成し，RGAT に加える手法を提案する．提案手法の距離の情報を図 4.3 に示す．

距離を加える従来手法に，絶対位置 [Vaswani et al., 2017] や相対位置 [Shaw et al., 2018] がある．図 4.3 に示すように，絶対位置はノード (発話) の順番に基づき，相対位置はエッジ (発話間の関係) すなわち発話から発話への距離に基づく．発話から発話へ有向グラフを結ぶ GNN では，相対位置の適用が望ましい．しかしながら，相対位置はエッジの種類によらず一律に距離の情報が与えられる．本手法は，自己依存と他者依存の取得を目的に RGAT を利用するため，エッジの種類に応じた位置の埋め込みを作成する．以上より，図 4.3 に示す，エッジの種類ごとに用意した相対位置を新たに作成する．

次に提案手法の式を示す．はじめに，隣接発話の集合の各要素に対応する位置の集合を定める．隣接発話の集合 $\mathcal{N}_r(i)$ に対応する位置の集合を，$\mathcal{I}_r(i) \subset \mathbb{N}$ とする．

[3]対象ノードと隣接するノード間の，エッジの値に関するソフトマックス (Softmax) 関数を edge softmax と呼ぶ．

$\mathbb{N}$ は自然数の集合を示す．$\mathcal{I}_r(i)$ を，窓幅 $p(=f)$ を最大とする自然数から降順に，隣接発話の集合 $\mathcal{N}_r(i)$ の要素数を取り出した集合と定める．例えば，図 4.3 の例で窓幅 $p=3$ の時，エッジの種類 $r=1$ で発話 x_4 に隣接する集合は，$\mathcal{N}_1(4)=\{1,2,4\}$ である．この時，窓幅 $p=3$ を最大とする自然数から降順に，隣接発話の集合 $\mathcal{N}_1(4)$ の要素数取り出した位置の集合は $\mathcal{I}_1(4)=\{1,2,3\}$ となる．次に，エッジの種類 $r=2$ で発話 x_4 に隣接する集合 $\mathcal{N}_2(4)=\{3\}$ に対応する位置の集合を考える．この時，窓幅 $p=3$ を最大とする自然数から降順に，隣接発話の集合 $\mathcal{N}_2(4)$ の要素数取り出した位置の集合は $\mathcal{I}_2(4)=\{3\}$ となる．

続いて，隣接発話の集合 $\mathcal{N}_r(i)$ の要素に対応する，位置の集合 $\mathcal{I}_r(i)$ の要素を取り出す変換を定める．隣接発話の集合 $\mathcal{N}_r(i)$ の中で，識別対象 i から相対位置の遠い順に，位置の集合 $\mathcal{I}_r(i)$ の要素を対応させる．その変換を $\mathrm{idx}_{r,i}:\mathcal{N}_r(i)\rightarrow\mathcal{I}_r(i)$ とする．すなわち，対象 $i=4$ にエッジの種類 $r=4$ で隣接する発話の集合 $\mathcal{N}_4(4)=\{5,9,10\}$ に対して，対象 $i=4$ から相対位置の遠い順に，位置の集合 $\mathcal{I}_4(4)=\{1,2,3\}$ を対応させる．例えば，$j=10$ 番目の発話は，隣接する発話 $\mathcal{N}_4(4)$ の中で対象 $i=4$ に最も遠いので，位置の集合 $\mathcal{I}_4(4)$ の中から 1 が対応する．したがって，$j=10$ 番目の発話の位置を抜き出すときは，変換 $\mathrm{idx}_{r,i}(j)$ を用いて $\mathrm{idx}_{4,4}(10)=1$ となる．

本手法では，発話 x_j から発話 x_i への位置の埋め込みを，変換 $\mathrm{idx}_{r,i}$ を用いて下式で表現する．

$$\mathrm{PE}_{ij}=\mathrm{idx}_{r,i}(j)-\{(r+1)\bmod 2\} \tag{4.3}$$

mod は剰余演算を示す．式 (4.3) の mod を含む 2 項目は，他者の依存関係を示すエッジの種類 $r=2,4$ の時に，位置の埋め込みを (-1) 加算する処理を示す．すなわち，対象の発話に対して，自分自身の発話の位置の埋め込みを最大 (窓幅 p) に設定する．また，位置の埋め込み PE_{ij} は，対象の発話から離れるに従い減少する．

位置埋め込み PE_{ij} の値は固定する場合と学習する場合の 2 つを用意する．まず値を固定する場合は，式 (4.3) によって得られる値をそのまま利用する．一方で，値を学習する場合は，式 (4.3) によって得られた値を入力した 1 層の FFN の出力を利用する．値を学習する場合は，識別対象の発話に適した動的な位置埋め込みの表現を学習することが可能になる．

図 4.4 に示すように，相対位置に基づく提案手法 PE_{ij} は，発話から発話への距離に相当し，エッジの重み係数に加えることができる．新たに提案する位置埋め込みを加えたエッジの重み係数を下式で示す．

$$\begin{aligned}\alpha_{ij}^{(l)}&=\frac{\exp(\beta_{ij}^{(l)})}{\sum_{\bar{j}\in\mathcal{N}_r(i)}\exp(\beta_{i\bar{j}}^{(l)})}\\ \beta_{ij}^{(l)}&=\mathrm{LeakyReLU}\big((\mathbf{a}_r^{(l)})^T[W_r^{(l)}\mathbf{h}_i^{(l)}||W_r^{(l)}\mathbf{h}_j^{(l)}]+\mathrm{PE}_{ij}\big)\end{aligned} \tag{4.4}$$

式 (4.2) との違いは，位置埋め込み PE_{ij} を足し合わせる点である．位置埋め込み PE_{ij} は，値を固定する場合も学習する場合も，同様にスカラー値である．

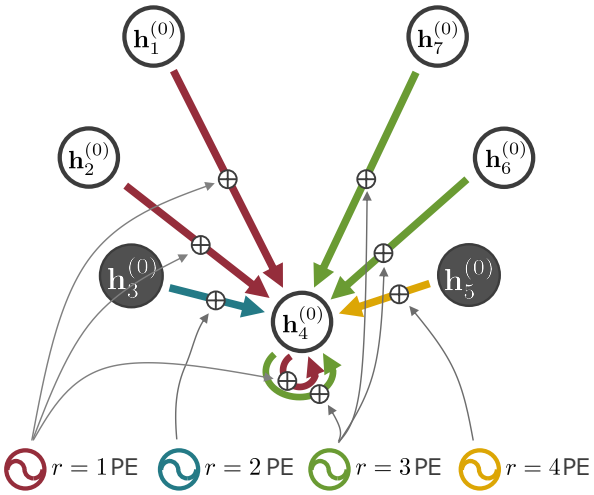

図 4.4: 提案手法の追加方法．位置埋め込みをエッジの種類ごとに用意し，それぞれをエッジに加える．“PE” は提案する位置埋め込みを示す．

4.2.3　感情ラベルの識別

4.2.2 項で発話間の関係を考慮した特徴量 $\mathbf{h}_i^{(L)}$ と，4.2.1 項で作成した発話の内容を示す特徴量 $\mathbf{h}_i^{(0)}$ を連結したベクトル $\mathbf{x}$ を入力し，ReLU (Rectified Linear Unit) 活性化関数を間に挿入した 2 層の FFN を用いて，感情ラベルを判別する．式を以下に示す．

$$\text{Classifier}(\mathbf{x}) = \text{ReLU}(\mathbf{x}W_1 + \mathbf{b}_1)W_2 + \mathbf{b}_2 \tag{4.5}$$

W_1 と W_2 は FFN のパラメータを示し，$\mathbf{b}_1$ と $\mathbf{b}_2$ は学習可能なバイアスベクトルを示す．

4.3　実験設定

4.3.1　従来手法

提案手法の有効性を検証するために，以下に示す従来手法と精度を比較する．

CNN [Kim, 2014]　CNN を用いて対話における各発話の特徴量を利用する手法である．これは対話における発話同士の関係は利用しない．

CNN+cLSTM [Poria et al., 2017]　CNN を用いた発話の特徴量ベクトル作成に加えて，双方向 LSTM を用いて隣接する発話間の関係を利用する手法である．

KET [Zhong et al., 2019] 階層的な自己注意層を用いて発話の特徴量を利用する手法である．また，GAT と外部データベースを組み合わせ，常識的知識と感情に関連する単語の特徴量を利用する手法である．

DialogueRNN [Majumder et al., 2019] CNN を用いて発話の特徴量ベクトルを作成し, 各発話の関連と話者の特徴, 感情の推移について, それぞれ GRU [Chung et al., 2014] でモデリングする手法である．

DialogueGCN [Ghosal et al., 2019] CNN を用いて発話の特徴量ベクトルを作成し，隣接する発話間の相互作用を GRU を用いて利用する手法である．加えて，自己依存と他者依存の取得に RGAT を利用する手法である．

IEIN [Lu et al., 2020] 双方向 GRU を用いて発話の特徴量ベクトルを作成し，発話間の相互作用の利用に双方向 GRU と注意機構を用いる手法である．さらに，感情ラベルの確率を入力し，新たな感情ラベルの確率を出力するネットワークを用いる．再起的に確率を導出することで，隣接する発話間の影響を利用する．

HiTrans [Li et al., 2020a] EmoryNLP データセットにおいて，高い認識性能を示した手法である．発話の特徴量ベクトル作成に BERT を用い，発話間の関係のモデリングに Transformer を利用し，階層的に組み合わせた手法である．

DialogXL [Shen et al., 2020] MELD, IEMOCAP データセットにおいて，高い認識性能を示した手法である．過去の発話を保存し共有するネットワークを，XLNet [Yang et al., 2019b] に加えた手法である．また隣接する発話間の関係 (局所的) と，会話全体の発話間の関係 (大域的) と，話し手，聞き手の特徴を，それぞれ自己注意層を用いて利用する手法である．

4.3.2 評価方法

表 3.1 に示す，対話の感情認識における 3 つのベンチマークセットを用いて，提案手法の有効性を検証する．Ghosal らの手法 [Ghosal et al., 2019] で用いられた評価指標と同じ，重み付き F1 値を全てのデータセットの評価に用いる．

4.3.3 モデルの学習

提案手法は, 損失関数に交差エントロピー (CE: Cross Entropy) 損失を用いて学習を行った．また, 学習率を $1e$-3 に設定し, Cosine Annealing Schedule により学習率を減少させ学習した [Loshchilov and Hutter, 2016]．RAdam optimizer [Liu et al.,

モデル	MELD	IEMOCAP	EmoryNLP
CNN	55.86	48.18	32.59
CNN+cLSTM	56.87	54.95	32.89
KET	58.18	59.56	34.39
DialogueRNN	57.03	62.75	31.70
DialogueGCN	58.10	64.18	-
IEIN	60.72	64.37	-
HiTrans	61.94	64.50	**36.75**
DialogXL	62.41	65.94	34.73
提案手法	**63.12** $\pm$ 0.65	**65.95** $\pm$ 1.92	35.58 $\pm$ 0.82

表 4.1: MELD, IEMOCAP, EmoryNLP ベンチマークデータセットにおける従来手法との比較．ボールド体は最も性能が高い値を示す．各値は 5 回の実験による重み付き F1 値の平均値を示す．

2019a] を用いて最適化を行った．また, RGAT のレイヤ数とバッチサイズはそれぞれ 2 と 8 に設定し, エッジ重み係数を計算する際に 4-Head Attention を用いた．ドロップアウトの割合は $(0.1, 0.3, 0.5)$ の中から, 窓幅は $(p, f) = (2, 2), (3, 3), (4, 4), (5, 5)$ の中から，検証セットで最も重み付き F1 値が高くなるものを選択した．全ての実験結果は 5 回行い平均値を用いた．

4.4　結果と考察

4.4.1　従来手法との比較

従来手法との比較結果を表 4.1 に示す．提案手法を除く全ての重み付き F1 値は, 各文献から引用する．

表 4.1 より，MELD データセットでは重み付き F1 値 63.12% を示し，従来手法を約 0.7% 近く上回る高い認識性能を示した．また，IEMOCAP データセットにおいても，重み付き F1 値 65.95% を示し，高い認識性能を示した．結果から 3 つのベンチマークデータセットのうち，2 つのデータセットで最高峰の認識性能を示し，提案手法の有効性を確認した．さらに，複数のベンチマークで高い認識精度を有することから，データ数や発話数，登場する話者数が異なる場合でも，精度良く認識できることを確認した．

♯	モデル	*neutral*	*surprise*	*fear*	*sadness*	*joy*	*disgust*	*anger*	W-F1
0	CNN	77.24	50.54	0.32	22.28	54.19	2.86	42.88	58.48
1	CNN+cLSTM	76.47	50.17	0.92	26.51	55.62	9.65	46.77	59.33
2	DialogueRNN	76.23	49.59	0.00	26.33	54.55	0.81	46.76	58.73
3	DialogueGCN	76.02	46.37	0.98	24.32	53.62	1.22	43.03	57.52
4	IEIN	77.52	53.65	3.31	23.62	56.63	19.38	48.88	60.72
5	RoBERTa	75.83 (± 1.37)	52.50 (± 1.89)	**22.87** (± 3.06)	36.92 (± 3.01)	56.10 (± 1.24)	27.38 (± 5.89)	45.18 (± 2.51)	60.85 (± 0.85)
6	+RGAT (RoBERTa+RGAT)	76.77 (± 0.64)	51.78 (± 0.96)	11.87 (± 3.02)	33.76 (± 3.13)	58.65 (± 1.42)	21.97 (± 2.70)	47.16 (± 2.41)	61.27 (± 0.69)
7	+PE (RoBERTa+RGAT+PE)	**77.55** (± 0.87)	**55.72** (± 1.80)	14.39 (± 5.69)	**37.60** (± 0.56)	**59.84** (± 0.83)	**27.51** (± 5.10)	**49.93** (± 1.61)	**63.12** (± 0.65)

表 4.2: MELD データセットにおける従来手法と提案手法の，感情ラベルごとの認識結果．ボールド体は最も性能が高い値を示す．RoBERTa は，RoBERTa を用いて発話の内容を示す特徴量ベクトルを作成する手法，+RGAT は，発話間の関係の利用を目的として RGAT を加えた手法 (RoBERTa+RGAT)，+PE は，提案する距離の情報を加えた手法 (RoBERTa+RGAT+PE) を示す．W-F1 は重み付き F1 値を示す．

4.4.2 感情ラベルの比較とアブレーション分析

MELD データセットを用いて，感情ラベルごとの認識性能を比較し，提案手法の特徴を分析する．さらに，提案手法における発話間の関係と距離の有効性を検証する．結果を表 4.2 に示す．従来手法 (♯0 − ♯4) に加え，♯5 に発話の内容を示す特徴量の作成を目的とした RoBERTa(RoBERTa)，♯6 に発話間の関係の利用を目的として RGAT を加えた手法 (+RGAT)，そして ♯7 に距離の情報を加えた提案手法 (+PE) の結果を示す．従来手法 (♯0 − ♯4) のラベルごとの F1 値と重み付き F1 値は，各文献から引用する．

結果から，従来手法 (♯0 − ♯4) と比較し，全ての感情ラベルで提案手法 (♯7) の有効性を確認できる．特に，表 3.3 に示す出現回数の少ない *disgust* ラベルにおいて，提案手法 (♯7) は F1 値 27.51% を示し，F1 値 19.38% を示す IEIN(♯4) を 8% 以上上回る認識性能を示した．出現頻度の低い感情ラベルにおいても，提案手法は高い認識性能を有することを確認した．

RoBERTa(♯5) と +RGAT(♯6) と比較し，提案手法 (♯7) は最も高い F1 値 63.12% を示し，その有効性を確認できる．また，ほとんど全ての感情ラベルにおいても，提案手法 (♯7) は高い認識性能を示す．しかしながら，*fear* ラベルでは，RoBERTa(♯5) が高い認識性能を達成した．MELD データセットでは，“I lost it” や “How bad is this?” といった発話に，しばしば *fear* ラベルが付与される．RoBERTa(♯5) が最も高い F1 値を示すことから，他の発話からの影響に比べて，識別対象の発話の内容が *fear* の感情に関連することが分かった．具体的な事例分析を，4.4.5 項で議論する．

♯	発話の内容	発話間の関係	距離の情報	タイプ	重み付き F1
0	✓	-	-	-	60.85 ± 0.85
1	✓	✓	-	-	61.27 ± 0.69
2	✓	✓	絶対位置	固定	62.45 ± 0.82
3	✓	✓		学習	61.89 ± 0.78
4	✓	✓	相対位置	固定	61.34 ± 1.23
5	✓	✓		学習	62.06 ± 0.70
6	✓	✓	提案手法	固定	62.41 ± 0.38
7	✓	✓		学習	**63.12 ± 0.65**

表 **4.3:** 絶対位置と相対位置，提案手法に基づく位置の比較.

4.4.3　位置埋め込みの比較

続いて, MELD データセットを用いて, 絶対位置と相対位置と比較し, 依存関係の種類ごとに位置の埋め込みを追加する提案手法の有効性を確認する．Transformer における位置埋め込み [Vaswani et al., 2017] を参考に，絶対位置に基づく位置埋め込みを RGAT の入力部分に加え，相対位置に基づく位置埋め込みをグラフのエッジ係数に加える．位置埋め込みを示す値は，固定する場合と，FFN を用いて学習する場合のいずれかを用いる．

結果を表 4.3 に示す．発話の内容を示す特徴量の作成を目的とした RoBERTa(♯0) と，発話間の関係の利用を目的として RGAT を加えた手法 (♯1) をベースラインとする．距離の種類を変更し，固定または学習した場合の結果を ♯2 − ♯7 に示す．

実験結果から，依存関係の種類に応じた位置埋め込みの値を学習する場合 (♯7)，最も高い重み付き F1 値 63.12% を得た．提案手法はベースライン (♯0, ♯1) に比べ 2% 近く値が改善し，さらに絶対位置や相対位置に比べ高い認識精度を示した．以上より，提案手法の有効性を確認できる．また，絶対位置や相対位置，提案手法 (♯2 − ♯7) が，ベースライン (♯0, ♯1) よりも高い重み付き F1 値を得ていることから，距離の有効性も確認することができる．

4.4.4　未来の依存関係の効果

次に，未来の依存関係の有効性を分析する．提案手法はデータが全て揃うオフラインへの応用を想定し，4.2.2 項に示す (1) 自己 - 過去， (2) 他者 - 過去， (3) 自己 - 未来， そして (4) 他者 - 未来の計 4 種類のエッジを設定した．しかし，対話の感情認識は，対話システムの構築 [Majumder et al., 2020] などのリアルタイム性が求められるアプリケーションへの応用も考えられる．このようなリアルタイムへの応用を想定した場合，計 4 種類の依存関係の中で未来の依存関係は利用

♯	モデル	重み付き F1
0	RoBERTa	60.85 ± 0.85
1	+RGAT (RoBERTa+RGAT)	61.27 ± 0.69
2	+PE (RoBERTa+RGAT+PE)	**63.12 ± 0.65**
3	−未来の関係	61.84 ± 0.61

表 4.4: 未来の依存関係を除きリアルタイム性を考慮した提案手法 (♯3) の認識性能．発話の内容を示す特徴量の作成を目的とした RoBERTa(♯0) と，発話間の関係の利用を目的として RGAT を加えた手法+RGAT(♯1)，未来の依存関係を含めた計 4 種類のエッジを利用する提案手法+PE(♯0) を比較.

できない．そこで，本実験は，未来の依存関係を除く (1) 自己 - 過去， (2) 他者 - 過去の計 2 種類のエッジのみを利用し，リアルタイム性を考慮する提案手法の認識性能を検証する．

結果を表 4.4 に示す．未来の依存関係を除いた手法 (♯3) の性能を検証するために，発話の内容を示す特徴量の作成を目的とした RoBERTa(♯0) と，発話間の関係の利用を目的として RGAT を加えた手法+RGAT(♯1)，未来の依存関係を含めた計 4 種類のエッジを利用する提案手法+PE(♯2) の結果も合わせて再掲し比較する．

実験結果から，未来の依存関係を除いた 2 種類の関係を利用する提案手法 (♯3) は重み付き F1 値 61.84% を示し，4 種類の依存関係を利用する提案手法 (♯2) に比べて，約 1.5% の重み付き F1 値が低下した．これは，後方（未来）の発話にも現れる，対象発話の感情を識別するための情報が利用できなくなったためである．以上の結果から，リアルタイム性が求められるアプリケーションに提案手法を応用する場合，性能の劣化を許容しなければならない．今後は，リアルタイムなアプリケーションへの応用も想定し，過去の発話を効果的に利用するグラフニューラルネットの構造を検討する．

4.4.5 事例分析

最後に，発話の内容を示す特徴量の作成を目的とした RoBERTa，発話間の関係の利用を目的として RGAT を加えた手法+RGAT，そして距離の情報を加えた提案手法+PE，それぞれの推定結果を分析する．また，エッジの種類に応じた発話間の関係を考慮する RGAT に代わり，エッジの種類を区別しない自己注意 (SA: Self Attention) 層を利用する手法+SA（RoBERTa+SA）を用意し，4 つの依存関係によってエッジの種類を区別する提案手法の有効性を確認する．さらに，+SA と+RGAT，提案手法+PE が，周辺のどの発話の影響を受けて推定したかを分析する．

結果を図 4.5 に示す．図 4.5 の左は，MELD の検証セットの一部で，7 番目の発

順番	話者	発話	ラベル	RoBERTa	+SA	+RGAT	+PE
0	Ross	Wow! It looks like we got a lot of good stuff.	*joy*	*surprise*	*surprise*	*joy*	*joy*
1	Rachel	Oh we did, but my mom got us the greatest gift of all.	*joy*	*joy*	*joy*	*joy*	*joy*
2	Ross	A	*neutral*	*neutral*	*neutral*	*neutral*	*neutral*
3	Rachel	No. She's going to live with us for eight weeks.	*neutral*	*neutral*	*neutral*	*neutral*	*joy*
4	Ross	Uh, what?	*surprise*	*neutral*	*surprise*	*surprise*	*neutral*
5	Rachel	Yes! She's gonna help us take care of the baby! Woo-hoo.	*joy*	*joy*	*joy*	*joy*	*joy*
6	Ross	What You're not serious.	*surprise*	*surprise*	*surprise*	*surprise*	*surprise*
7	Ross	I mean she's a very nice woman, but there is no way we can take eight weeks of her.	*anger*	*neutral*	*sadness*	*joy*	*anger*
8	Ross	She'll drive us totally crazy.	*disgust*	*anger*	*disgust*	*joy*	*anger*
9	Mrs. Green	Hi Ross!	*joy*	*joy*	*joy*	*joy*	*joy*
10	Ross	Hi roomie!	*joy*	*joy*	*joy*	*joy*	*joy*

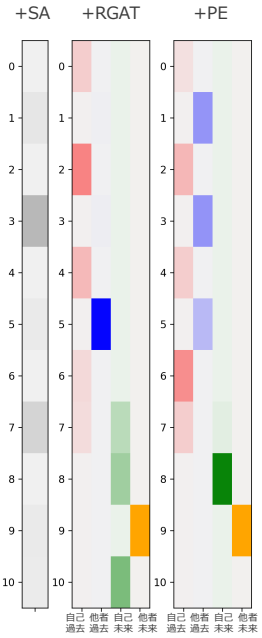

図 4.5: MELD の検証セットの一部と，+SA と+RGAT，+PE のエッジ重み係数．左の表は，MELD の検証セットの一部で，義母の長期滞在について夫が戸惑いを感じるシーンを示す．右の図は，7 番目の発話を識別対象としたときの，発話間の関係の利用を目的として自己注意層を加えた手法+SA(RoBERTa+SA) と発話間の関係の利用を目的として RGAT を加えた手法+RGAT(RoBERTa+RGAT)，距離の情報を加えた提案手法+PE(RoBERTa+RGAT+PE) のエッジの重み係数を示す．重み係数の値が大きいほど，濃い色を示す．

話で RoBERTa と+SA，+RGAT が誤って識別し，提案手法+PE が正しく識別した例を示す．図 4.5 の左の表は，義母の長期滞在について夫が戸惑いを感じるシーンを示す．順番と話者，発話の内容，正解ラベル，RoBERTa と+SA と+RGAT と+PE の推定結果によって構成される．図 4.5 の右は，7 番目の発話を識別対象としたときの，+SA の注意重みと式 (4.2) に示す+RGAT のエッジ重み係数，式 (4.4) に示す+PE のエッジの重み係数を表す．4 種類の依存関係ごとに色を分けて可視化し，重み係数の値が大きいほど濃い色を表す．

図 4.5 の左表の結果が示すように， 自己注意層によって発話間の関係を利用する+SA は，7 番目の発話を誤って *sadness* と識別した．左表の一連の会話から，レイチェルの母親が長期滞在する計画を知ったロスの 7 番目の発話は，ネガティブな感情を示す．さらに，直前の 6 番目の発話と直後の 8 番目の発話にロスの戸惑いが現れる．しかし，図 4.5 の右の結果から，+SA の重み係数は 3 番目の発話の影響が大きいことがわかる．すなわち，7 番目の発話に関連する自身の発話（6 番目の発話など）に注目せずに，関連のない 3 番目のレイチェルの発話に注目してしまい誤って識別した．

また，発話間の関係を利用する+RGAT は，7 番目の発話を誤って *joy* と識別し

順番	話者	発話	ラベル	RoBERTa	+SA	+RGAT	+PE
0	Rachel	God. I forgot how much I love driving. I have	*joy*	*joy*	*sadness*	*joy*	*joy*
1	Ross	You don't have a	*neutral*	*surprise*	*neutral*	*neutral*	*surprise*
2	Rachel	Oh Ross you're so tense! You just gotta relax, okay? Just need to relax all right? Just need to relax	*neutral*	*neutral*	*sadness*	*anger*	*fear*
3	Ross	What-what are you doing?! Are you Okay that's not funny! Just stop horsing around!	*anger*	*anger*	*anger*	*anger*	*disgust*
4	Rachel	I am not horsing around, okay? I am Porsching around.	*joy*	*anger*	*anger*	*anger*	*anger*
5	Rachel	Uh-oh.	*fear*	*fear*	*neutral*	*neutral*	*neutral*
6	Ross	Okay, stay calm. Nothing is going to happen to you, you are not in that much trouble.	*neutral*	*neutral*	*neutral*	*neutral*	*neutral*
7	Rachel	Really? You think so?	*fear*	*surprise*	*surprise*	*surprise*	*surprise*
8	Ross	I was talking to myself! You're going down!	*anger*	*anger*	*anger*	*anger*	*anger*

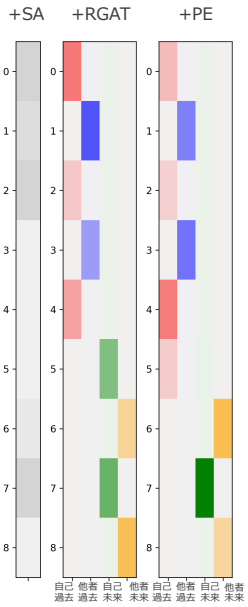

図 4.6: MELD の検証セットの一部と，+SA と+RGAT，+PE のエッジ重み係数．左の表は，MELD の検証セットの一部で，楽観的なレイチェルの運転のせいでロスがパニックに陥るシーンを示す．右の図は，5 番目の発話を識別対象としたときの，発話間の関係の利用を目的として自己注意層を加えた手法+SA(RoBERTa+SA) と発話間の関係の利用を目的として RGAT を加えた手法+RGAT(RoBERTa+RGAT)，距離の情報を加えた提案手法+PE(RoBERTa+RGAT+PE) のエッジの重み係数を示す．重み係数の値が大きいほど，濃い色を示す．

た．図 4.5 の右の結果から，+RGAT の重み係数は赤色の自己 - 過去の関係 $(r = 1)$ の中で関連のない 2 番目の発話の影響が大きく，また青色で示す他者 - 過去の関係 $(r = 2)$ の中で 5 番目のレイチェルの発話の影響が大きいことが分かる．すなわち，+RGAT は 7 番目の発話に関連する自身の発話（6 番目の発話など）に注目せずに，*joy* と認識する 5 番目のレイチェルの発話に共感してしまい，7 番目の発話を誤って識別した．

一方で，距離の情報を加えた提案手法+PE は，左表の 7 番目の発話を *anger* と正しく認識した．右図が示すように+PE の重み係数は，赤色の自己 - 過去の関係 $(r = 1)$ の中で直前の 6 番目の発話の影響が+RGAT に比べて大きく，緑色の自己 - 未来の関係 $(r = 3)$ の中で直後の 8 番目の発話の影響が大きくなった．このことから，+PE は距離の情報を加えることで，ロスの戸惑いが現れる前後の発話に注目し，結果的に 7 番目の発話を正しく認識した．

次に，提案手法+PE が誤って識別した例を図 4.6 に示す．図 4.6 の左は，MELD の検証セットの一部で，4.4.2 項の表 4.2 の結果で，*fear* の感情を示す発話を RoBERTa が正しく識別し，+RGAT，提案手法+PE が誤って識別した例を示す．図 4.6 の左の表は，楽観的なレイチェルの運転のせいでロスがパニックに陥るシーンを示す．図 4.6 の右は，+SA と+RGAT，+PE のそれぞれが 5 番目の発話を識別する

際に，周辺のどの発話の影響を重視したかを可視化する．

図 4.6 の左表の結果が示すように，発話間の関係を利用する+SA と+RGAT，提案手法+PE は，5 番目の発話を誤って *neutral* と識別した．左表の一連の会話から，車の制御を失ってしまったレイチェルの 5 番目の発話は，*fear* の感情を示す．しかし，図 4.6 の右の結果から，+SA は，識別対象の 5 番目の発話の影響が小さいことがわかる．また，+RGAT と提案手法+PE の両方も，赤色の自己 - 過去の関係 ($r = 1$) と緑色の自己 - 未来の関係 ($r = 3$) の中で，識別対象の 5 番目の発話の影響が小さいことが分かる．すなわち，車の制御を失った状況を示す自身の発話に注目しなかった．識別対象の発話の内容を利用する RoBERTa は正しく認識していることから，個々の発話の依存関係を利用するモデルが悪影響を与えてしまったことが分かる．この例を正しく識別するためには，レイチェルが置かれている状況を把握する必要があり，発話の内容を適切に把握する必要がある．

4.5　発話の距離に基づく発話間の関係を利用する手法のまとめ

本章は，対話における各発話の感情認識において，RGAT に適した距離の情報を加える方法を提案した．距離の情報を加えた RGAT を用いることで，発話間の関係と発話の距離の両方を利用できる．3 つのベンチマークデータセットを用いて提案手法の有効性を確認したところ，2 つのデータセットで従来手法を上回る認識性能を示した．また，評価実験を通して，依存関係の種類に応じた距離の情報を用いた手法が高い認識性能を示し，有効性を確認した．

第5章　会話の履歴と発話間の関係の組み合わせ

5.1　研究の概要

本章は，会話の履歴を利用する識別モデルと発話間の関係を利用する識別モデルを組み合わせる手法を提案する．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いもの（近傍事例）を検索し，動的な重み付き線形和によって補強する事例ベース手法を提案する．

対話の感情認識では，1.2.3 項に示すように，発話の内容だけでなく発話間の関係が話者の感情に大きな影響を与えることが知られているため [Poria et al., 2019]，4.2 節では発話の距離と発話間の関係を利用する手法を提案した．また，近年では，発話間の関係を効果的に利用する手法が提案されている [Shen et al., 2021]．しかしながら，これらの手法は一連の会話の履歴の利用が容易でない．表 1.2 と表 1.3 を用いて，会話の履歴の重要性を示す．たとえば，表 1.2 の 7 番目の発話 “Yes” は，それまでの発話の内容から負の感情を示すが，表 1.3 の 5 番目の発話 “Yes” は正の感情を示す．このように，同じ発話であっても，一連の会話の履歴に応じて異なる感情を示すことがある．発話間の関係を考慮する従来手法 [Ghosal et al., 2019, Shen et al., 2021] や 4.2 節に示す提案手法は，識別対象の発話と周辺の発話との対の関係を利用することは可能だが，会話全体の履歴を把握することは容易でない．

会話の履歴を利用する代表的な方法として，連続した複数の発話を連結し，事前学習済み BERT に入力する方法 [Yang et al., 2019a] がある．この手法は，識別対象の発話とその先行文脈を言語モデルに入力し，会話全体に注意を向けるため，会話の履歴を利用することが可能である．しかしながら，この手法は会話全体に注意を向けるため，逆に個々の発話の依存関係の利用が難しい．1.2.3 項に示すように，話者の感情は発話間の関係や会話の履歴に依存するため，どちらの影響も重要である．

そこで，本章は発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組み合わせるアンサンブル手法を提案する．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いものを検索し，動的な重み付き線形和によって補強する事例ベース手法を提案する．具体的には，まず識別対象の発話とその先行文脈をクエリーとして，会話の履歴の観点で意味的に近い発話を訓練データセットから k 近傍法を用いて検索する．検索した発話 (近傍事例) に付与された感情ラ

ベルと，識別対象の発話との距離を基に感情ラベルの確率分布を作成し，発話間の関係を利用するモデルの確率分布と重み付き線形和によって組み合わせる．提案手法を用いることで，発話間の関係と会話の履歴の両方の特徴を利用することができる．

さらに，定数による重み付き線形和で 2 つの確率分布を組み合わせるだけでなく，識別対象の発話ごとに動的に重み係数を変更する方法を提案する．定数による重み係数は，常に一定の割合で近傍事例による確率分布を利用するため，近傍事例に適切な事例が存在するか否かに応じて重み係数を調整することができない．そこで本章は，識別対象の発話に応じて，動的に重み係数を変更する方法を提案する．具体的には，発話間の関係を利用するモデルから得られる識別対象発話の特徴量と近傍事例の特徴量を入力し，重み係数を導出するニューラルネットワークを構築する．重み係数のネットワークを学習するために，発話間の関係を利用するモデルによる確率分布と近傍事例による確率分布のそれぞれが示す感情ラベルが，教師ラベルと一致する場合に重み係数を高め，そうでない場合に重み係数を低くする損失関数 (係数損失) を導入する．

ERC における 3 つのベンチマークデータセットによる評価実験を通して，動的に重み係数を変更する提案手法が最高水準の認識性能を示し，有効性を確認した．加えて，重み係数の頻度分布を検証する実験を通して，適切な重み係数を学習するためには，係数損失が必要であることを確認した．

5.2　先行発話と会話の履歴を利用するモデルの識別性能

本論文は，会話の履歴を利用する識別モデルと発話間の関係を利用する識別モデルを組み合わせる手法を提案する．提案手法は，会話の履歴を利用するモデルとして，Yang らの手法 [Yang et al., 2019a] を参考に，識別対象の発話とその先行文脈を連結し入力する RoBERTa [Liu et al., 2019b] を用いる．この方法を用いることで，識別対象の発話とその先行文脈に注意を向けるため，会話の履歴を利用することができる．しかしながら，RoBERTa を含む言語モデルは入力系列の関連する部分に注意を向けるため，先行文脈を入力しても別の部分系列（例えば識別対象の発話）に注意が向き，先行文脈が利用されない可能性がある．

そこで，RoBERTa が会話の履歴を有効に利用しているかどうかを検証する．図 5.1 に，RoBERTa に入力する先行発話の数を $0, 1, 2, 3, 4, 5, 7, 10, 15, 20$ と変化させたときの感情認識の性能 (重み付き F1 値) の違いを示す．図 5.1 は，3.2 節に示す 3 つのベンチマークデータセットの検証セットにおいて，5 回実験を行った重み付き F1 値の平均値を示す．検証に使用する RoBERTa は，識別対象の発話と複数の先行発話を連結した対話を入力し，[CLS] トークンの位置の特徴量ベクトルを出力する．例えば，先行発話の数が 0 の場合は n 番目の識別対象の発話のみを入力し，先行発話の数が 3 の場合は n 番目の識別対象の発話と $n-1, n-2, n-3$ 番目の計 3

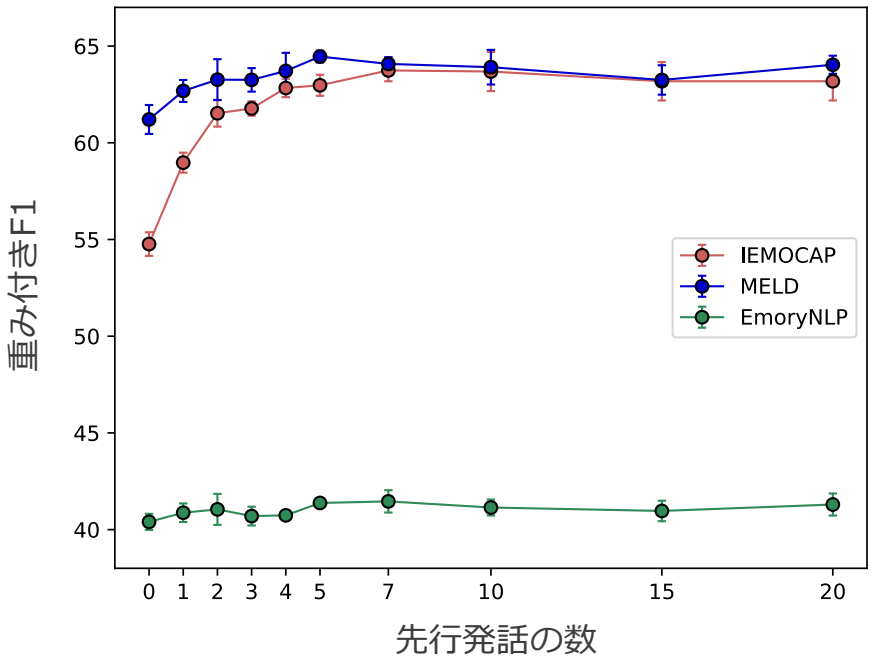

図 5.1: RoBERTa に入力する先行発話の数を変化させた時の認識性能．3 つのベンチマークデータセットの検証セットにおいて，5 回実験を行った重み付き F1 値の平均値．

つの先行発話を入力する．ただし，Yang らの手法 [Yang et al., 2019a] を参考に，発話間には [SEP] トークンを挿入する．また，GPU メモリの不足を回避するために，識別対象の発話とその発話から総トークン数が 128 を越えない範囲で遡った先行文脈を入力トークン列とする．得られた [CLS] トークンの位置の特徴量ベクトルから 2 層の FFN と Softmax 関数を用いて感情ラベルを出力する．RoBERTa は，先行発話の数ごとに，対話の感情認識ベンチマークの訓練セットでファインチューニングを行う．

図 5.1 の結果から，どのベンチマークデータセットにおいても，先行発話の数が 5 ～ 7 周辺の場合に，重み付き F1 値が高い値を示すことが分かる．先行発話の数が 0 や 1 の場合と比較して，5 ～ 7 周辺の場合に高い重み付き F1 値を示すことから，先行文脈を入力することで性能が向上することが分かる．以上の結果から，入力された会話の履歴は RoBERTa によって利用され，認識性能の向上に必要であることが分かった．本論文は，識別対象の発話とその先行文脈を連結し入力する RoBERTa を，会話の履歴を利用するモデルとして利用する．詳細を 5.4.2 項に示す．

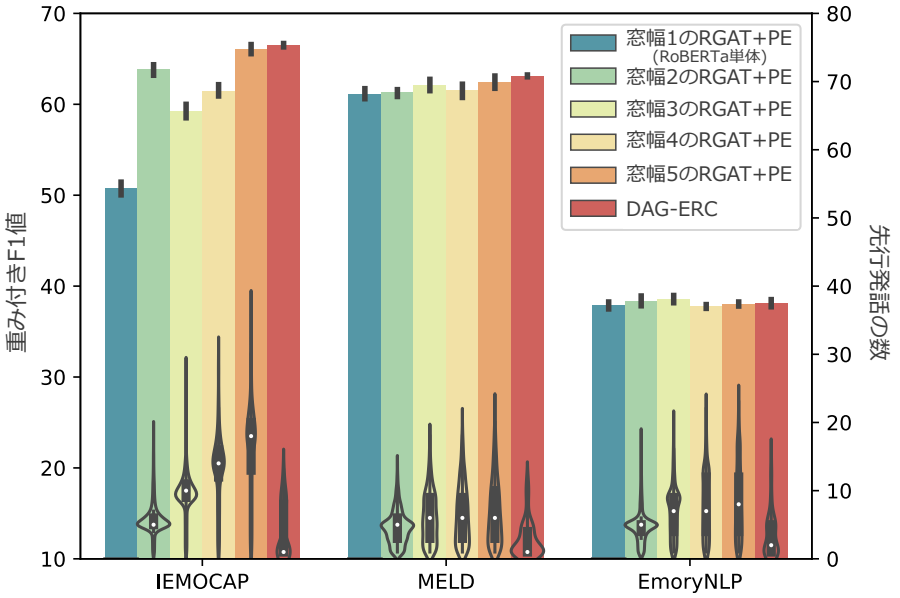

図 5.2: 先行発話の数と発話間の関係を利用するモデルの識別性能．感情認識の性能（重み付き F1 値）を棒グラフ（左軸）を用いて，識別モデルに入力する先行発話の数をバイオリン図（右軸）を用いて示す．

5.3　先行発話と発話間の関係を利用するモデルの性能

本論文は，会話の履歴を利用する識別モデルと発話間の関係を利用する識別モデルを組み合わせる手法を提案する．本章の提案手法は，発話間の関係を利用するモデルとして，有向非巡回グラフニューラルネットワークを用いて自己依存と他者依存の関係を利用し高い認識性能を示す DAG-ERC [Shen et al., 2021] を用いる．本節は，4 章に示す発話の距離を利用する提案手法と DAG-ERC に入力する隣接発話の数を計測し，その数と認識性能を比較することで，DAG-ERC の有効性を確認する．

図 5.2 に，窓幅（4.3.3 項）を変化させた時の発話の距離を利用する提案手法（RGAT+PE）と，DAG-ERC の感情認識の性能（重み付き F1 値）を示す．図 5.2 は，3.2 節に示す 3 つのベンチマークデータセットの検証セットにおいて，5 回実験を行った重み付き F1 値の平均値を示す．ただし，提案手法（RGAT+PE）の窓幅 1 は，識別対象の発話を利用する RoBERTa 単体の性能を示す．また，各ベンチマークの検証セットに対して，各識別モデルに入力する先行発話の数を計測し，その分布を示す．感情認識の性能を棒グラフ（左軸）で，識別モデルに入力する先行発話の数をバイオリン図（右軸）で可視化する．

図 5.2 の結果から，IEMOCAP と MELD では，窓幅 5 の場合に高い性能を示す提案手法（RGAT+PE）に比べて，DAG-ERC はさらに高い重み付き F1 値を示すことが分かる．EmoryNLP では，窓幅 3 の場合に高い性能を示す提案手法（RGAT+PE）と，DAG-ERC は同程度の値を示すことが分かる．また，DAG-ERC における先

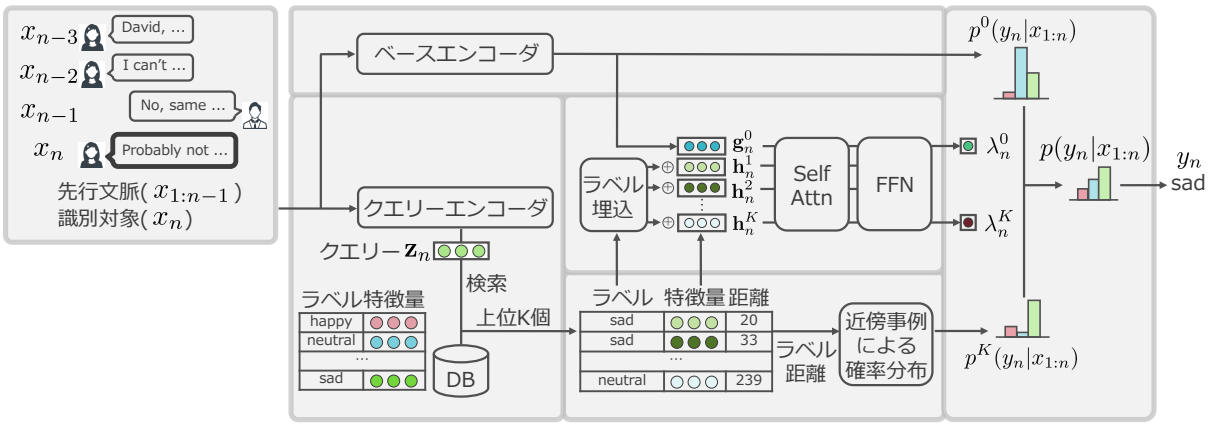

図 5.3: 提案手法の推論の流れ．提案手法は，ベースエンコーダ (発話間の関係を利用するモデル) による確率分布作成 (5.4.4 項)，近傍事例の検索 (5.4.5 項)，近傍事例による確率分布作成 (5.4.6 項)，確率分布の組み合わせ (5.4.7 項) で構成される．

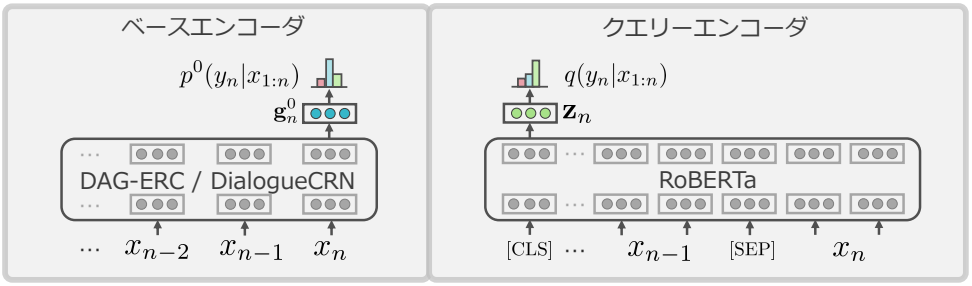

図 5.4: ベースエンコーダとクエリーエンコーダの構成

行発話の数は，概ね 10 未満の値に分布し，提案手法（RGAT+PE）よりも短い先行発話を利用することが分かる．以上の結果から，DAG-ERC は，短い先行発話から対象発話を識別するための情報を効果的に利用することが分かる．本論文は，より短い先行発話から効果的に情報を取得する DAG-ERC を，発話間の関係を利用するモデルとして利用する．詳細を 5.4.1 項に示す．

5.4 提案手法

提案手法の推論の流れを図 5.3 に示す．はじめに識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を発話間の関係を利用するモデル (ベースエンコーダ) に入力し，感情ラベルの確率分布 p^0 を作成する (5.4.4 項)．次に，識別対象の発話と先行文脈による対話を入力し，会話の履歴を利用するモデル (クエリーエンコーダ) を用いて得られる特徴量ベクトルをクエリーとして，あらかじめ訓練セットを用いて作成したデータベース (5.4.3 項) から意味的に近い事例を検索する (5.4.5 項)．ベースエンコーダとクエリーエンコーダの構成を図 5.4 に示す．検索した近傍事例に付与された感情ラベルと，識別対象の発話との距離に基づいて確率分布 p^K を作成

する (5.4.6 項)．さらに，ベースエンコーダから得られる識別対象発話の特徴量ベクトルと近傍事例の特徴量ベクトルを自己注意層 (SA: Self Attention) と FFN に入力し，重み係数 λ_n^0, λ_n^K を取得する．重み係数 λ_n^0, λ_n^K を用いて，ベースエンコーダの確率分布 p^0 と近傍事例の確率分布 p^K の重み付き線形和から確率分布 p を得る (5.4.7 項)．最後に，最も確率の高い感情ラベル y_n を出力する．

提案手法は，4 つのステップを通して，パラメータを段階的に学習し，感情ラベルの推論を行う．まずステップ 1 で，図 5.4 に示すベースエンコーダとクエリーエンコーダのパラメータを学習する．ベースエンコーダとクエリーエンコーダのそれぞれが出力する確率分布から得られる感情ラベルと教師ラベルを用いて，交差エントロピー (CE: Cross Entropy) 損失を計算する (5.4.1 項， 5.4.2 項)．次にステップ 2 では，ステップ 1 で学習したクエリーエンコーダを利用し，近傍事例検索用のデータベースを作成する (5.4.3 項)．ステップ 3 では，ステップ 1 で学習したベースエンコーダとクエリーエンコーダのパラメータを固定し，図 5.3 の重み係数 λ_n^0, λ_n^K を導出するネットワークのパラメータを学習する (5.4.7 項)．最後に，学習した全てのパラメータを固定し，図 5.3 の感情ラベルの推論を行う．

5.4.1　ベースエンコーダの学習

識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を入力し，n 番目の発話における感情ラベル y_n を導く確率分布を用いて，発話間の関係を利用するモデル (ベースエンコーダ) を学習する．図 5.4 に示すように，識別対象の発話と先行文脈による対話 $x_{1:n}$ を入力し，対象の n 番目の発話の特徴量ベクトル $\mathbf{g}_n^0$ を取得する．得られた特徴量ベクトル $\mathbf{g}_n^0$ から 2 層の FFN とソフトマックス (Softmax) 関数を用いて，確率分布 $p^0(y_n|x_{1:n})$ を作成する．

ベースエンコーダには，対話の状況に応じた相互作用と自身の発言による自己依存を利用する DialogueCRN [Hu et al., 2021]，話者自身と他者の発話の影響を考慮する DAG-ERC [Shen et al., 2021] のいずれかを用いる．DialogueCRN は，対話における各発話の内容を示す特徴量ベクトルを作成し，対話の状況に応じた相互作用と自身の発言による自己依存を利用する．状況に応じた相互作用を表す特徴量ベクトルと，自己依存を示す特徴量ベクトルを結合し，FFN と Softmax 関数を用いて感情ラベルの確率分布を作成する．

DAG-ERC は，RoBERTa-large [Liu et al., 2019b] を用いて対話における発話の内容を表す特徴量ベクトルを作成する．さらに，GAT を拡張した有向非巡回グラフニューラルネットワークを用いて自身の発話からの影響と他者の発話からの影響を考慮する特徴量ベクトルを取得する．発話の内容を表す特徴量ベクトルと発話間の影響を表す特徴量ベクトルを結合し，FFN と Softmax 関数を用いて感情ラベルの確率分布を作成する．ただし，DAG-ERC は，利用する先行発話の数が DialogueCRN と異なり，n 番目の発話の話者が直前に発した s $(s \leq n)$ 番目の発話 x_s から n 番目までの対話 $x_{s:n}$ を利用する．

ベースエンコーダのパラメータは，確率分布 p^0 が出力する感情ラベルと教師ラベルとの交差エントロピー損失を用いて学習する (ステップ 1)．ステップ 2 以降では，ステップ 1 で学習したベースエンコーダのパラメータを固定する．ベースエンコーダから得られる特徴量ベクトル $\mathbf{g}_n^0$ を図 5.3 の重み係数の導出に利用し (5.4.7 項)，確率分布 $p^0(y_n|x_{1:n})$ を図 5.3 の発話間の関係を利用するモデルによる確率分布として利用する (5.4.4 項)．

5.4.2 クエリーエンコーダの学習

次に識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を入力し，5.4.1 項と同様に n 番目の発話における感情ラベル y_n を導く確率分布を用いて，会話の履歴を利用するクエリーエンコーダを学習する．まず識別対象の発話とその発話から総トークン数が 128 を越えない範囲で先行文脈を遡ったトークン列を事前学習済 RoBERTa[1] に入力し，図 5.4 に示す [CLS] トークンの位置の特徴量ベクトル $\mathbf{z}_n$ を出力する．ただし，Yang らの手法 [Yang et al., 2019a] を参考に，発話間には [SEP] トークンを挿入する．得られた特徴量ベクトル $\mathbf{z}_n$ から 2 層の FFN と Softmax 関数を用いて，確率分布 $q(y_n|x_{1:n})$ を作成する．

クエリーエンコーダのパラメータは，確率分布 q が出力する感情ラベルと教師ラベルとの交差エントロピー損失を用いて学習する (ステップ 1)．ベースエンコーダと同様に，ステップ 2 以降では，ステップ 1 で学習したクエリーエンコーダのパラメータを固定する．クエリーエンコーダから得られる特徴量ベクトル $\mathbf{z}_n$ を，データベースの作成 (5.4.3 項) と近傍事例の検索 (5.4.5 項) に利用する．

5.4.3 データベース作成

訓練データセットの対話の内，識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を入力し，5.4.2 項で学習済みのクエリーエンコーダを用いて特徴量ベクトル $\mathbf{z}_n$ を作成する．対象の発話と先行文脈を入力し得られる特徴量ベクトルをキー，対象の発話に付与された感情ラベルを値として，訓練データセットの全ての対話の全発話から得られるキーと値の組みをデータベースに登録する (ステップ 2)．なお，本論文ではクエリーエンコーダとして，RoBERTa を用いるが，対象の発話とその先行文脈を入力し対象の発話の特徴量ベクトルを出力する他のモデルも利用可能である．

5.4.4 ベースエンコーダによる確率分布

図 5.3 に示す識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を入力し，n 番目の発話における感情ラベル y_n を導く確率分布 $p^0(y_n|x_{1:n})$ を，発話間の関係を利用するモデ

[1] https://huggingface.co/roberta-large

ル (ベースエンコーダ) を用いて作成する．ベースエンコーダは，5.4.1 項で学習済みの DialogueCRN [Hu et al., 2021] と DAG-ERC [Shen et al., 2021] のいずれかを用いる．なお，本論文ではベースエンコーダとして，DialogueCRN と DAG-ERC を用いるが，ERC タスクで学習した他のモデルも利用可能である．

5.4.5　近傍事例の検索

5.4.3 項で作成したデータベースを用いて，近傍事例を検索する方法について述べる．識別対象の発話 x_n とその先行文脈 $x_{1:n-1}$ を，5.4.2 項で学習済みのクエリーエンコーダに入力し，特徴量ベクトル $\mathbf{z}_n$ を取得する．次に，取得した特徴量ベクトル $\mathbf{z}_n$ を検索クエリーとして，5.4.3 項に示すデータベースに登録された特徴量ベクトルとの距離を計算し，距離の近い上位 K 個の近傍事例を取得する．取得した各近傍事例は，特徴量ベクトル，感情ラベル，クエリーとの距離によって構成され，その集合を $\mathbb{K}_n = \{(\mathbf{h}_n^k, y_n^{k\prime}, d(\mathbf{h}_n^k, \mathbf{z}_n)), k \in \{1, 2, \cdots, K\}\}$ とする．$\mathbf{h}_n^k$ は検索クエリー $\mathbf{z}_n$ を用いて 5.4.3 項のデータベースから取り出した近傍事例の特徴量ベクトルを示し，$y_n^{k\prime}$ は特徴量ベクトル $\mathbf{h}_n^k$ とデータベースの組みを形成する感情ラベルを示す．Khandelwal らの手法 [Khandelwal et al., 2021] を参考に，距離尺度 $d(\cdot, \cdot)$ として L2 ノルムを用いる．本論文では，訓練データセットの全ての対話を用いて検索用のデータベースを構築するが，一般的にデータベースのサイズが訓練セットよりも小さい場合，対象の発話に近い事例が少なくなり，ノイズとなる事例が検索される可能性がある．一方でサイズが訓練セットよりも大きい場合，事例の探索に時間がかかる．

さらに本稿は，Jiang らの手法 [Jiang et al., 2021] を参考に，訓練時と評価時で検索する事例の数 K を変更する．5.4.3 項で示すように，データベースは訓練データから構築するため，訓練時すなわち訓練データの発話をクエリーとする場合，完全に一致する発話をデータベースから検索する．一方で評価時すなわち評価データの発話をクエリーとする場合，類似した発話がデータベースに存在しない可能性がある．このような訓練時と評価時の違いは過学習を引き起こすことが知られているため [Jiang et al., 2021]，本稿は Jiang らの手法 [Jiang et al., 2021] と同様に，訓練時は上位 $K + 1$ 個の事例を検索し，最も類似する事例を削除した残りの K 個の事例を利用する．評価時は事例の削除を行わず，上位 K 個の事例を利用する．

5.4.6　近傍事例による確率分布

5.4.5 項で検索した近傍事例を用いて確率分布を作成する．検索した K 個の近傍事例の，感情ラベル $y_n^{k\prime}$ とクエリーとの距離 $d(\mathbf{h}_n^k, \mathbf{z}_n)$ を用いる．Khandelwal らの手法 [Khandelwal et al., 2021] を参考に，感情ラベル $y_n^{k\prime}$ と距離 $d(\mathbf{h}_n^k, \mathbf{z}_n)$ を用い

て，n 番目の発話の確率分布を算出する式を式 (5.1) に示す．

$$p^K(y_n|x_{1:n}) \propto \sum_{(\mathbf{h}_n^k, y_n^{k\prime})\in\mathbb{K}_n} \mathbb{1}_{y_n=y_n^{k\prime}} \exp\left(\frac{-d(\mathbf{h}_n^k, \mathbf{z}_n)}{T}\right) \tag{5.1}$$

$\mathbb{1}_{y_n=y_n^{k\prime}}$ は，近傍事例の感情ラベル $y_n^{k\prime}$ が感情ラベル y_n と同一である場合に 1 を返す指示関数である．T は距離の近い事例のラベルを重要視するか，頻度が多いラベルを重要視するかのバランスをとるハイパーパラメータである．T が小さい場合，距離の近い事例に付与された感情ラベルに重きが置かれ，そのラベルの確率が高い分布が作成される．一方 T が大きい場合，近傍事例に占める感情ラベルの出現頻度に重きが置かれ，出現頻度の高いラベルの確率が高い分布が作成される．

5.4.7 確率分布の組み合わせ

5.4.4 項で作成したベースエンコーダによる確率分布 p^0 と，5.4.6 項で作成した近傍事例による確率分布 p^K を組み合わせる．本論文は，定数の重み係数を導入し線形結合を行う方法と，動的に変化する重み係数を導入し線形結合を行う方法の 2 通りを示す．定数による重み係数は，常に一定の割合で近傍事例による確率分布を利用するため，近傍事例に適切な事例が存在するか否かに応じて重み係数を調整することができない．そこで，識別対象の発話に応じて動的に重み係数を変更する方法を示す．最後に，重み付き線形和によって得られる確率分布の中で，最も確率の高い感情ラベルを識別結果として出力する．なお，提案手法はベースエンコーダ (DAG-ERC または DialogueCRN) から得られる確率分布 p^0 と，クエリーエンコーダ (RoBERTa) を用いて検索した近傍事例による確率分布 p^K を組み合わせるため，異なるモデルの出力を組み合わせるアンサンブルの一種である．

静的な重み係数

重み係数 λ を用いて線形結合を行い，ベースエンコーダによる確率分布 p^0 と近傍事例による確率分布 p^K を式 (5.2) のように組み合わせる．

$$p(y_n|x_{1:n}) = (1-\lambda)\ p^0(y_n|x_{1:n}) + \lambda\ p^K(y_n|x_{1:n}) \tag{5.2}$$

重み係数 λ は定数で与えるハイパーパラメータである．常に一定の割合で 2 つの確率分布を組み合わせるため，本手法を提案手法 (静的な重み係数) とする．

式 (5.2) によって得られた確率分布の中で, 最も確率の高い感情ラベルを, 式 (5.3) に示すように識別結果として出力する．

$$\hat{y}_n = \arg\max_{c\in\mathbb{Y}} \left(p\left[c\right]\right) \tag{5.3}$$

ただし，$\mathbb{Y}$ は感情ラベルの集合，p は式 (5.2) で組み合わせた確率分布 $p(y_n|x_{1:n})$ を示す．

動的な重み係数

重み係数を導出するネットワークの構造　ベースエンコーダによる確率分布 p^0 が適切な場合はベースエンコーダ側の重み係数を高く，近傍事例による確率分布 p^K が適切な場合は近傍事例側の重み係数を高くすることが望まれる．どちらの確率分布が適切かは，ベースエンコーダから得られる識別対象発話の特徴量ベクトルと近傍事例の特徴量ベクトルの性質によって決まると考え，ベースエンコーダから得られる特徴量ベクトル $\mathbf{g}_n^0$ と，近傍事例の特徴量ベクトル $\mathbf{h}_n^k, k \in \{1, 2, \cdots, K\}$ を用いて重み係数を導出する．

次に，ラベル埋め込み表現を入力に加算する．5.4.5 項に示すように，近傍事例は訓練データセットから検索するため，感情ラベルが付与されている．近傍事例に付与された感情ラベルは，重み係数を決める重要な情報と考え，感情ラベルの埋め込み表現を近傍事例の特徴量ベクトル $\mathbf{h}_n^k, k \in \{1, 2, \cdots, K\}$ に加算する．ラベル埋め込みの種類は，感情ラベルのラベル数 C である．一方で，ベースエンコーダから得られる特徴量ベクトル $\mathbf{g}_n^0$ は，識別対象の発話を入力し作成するため，感情ラベルが与えられない．そこで，C 種類の感情ラベルとは異なるダミーラベルの埋め込み表現を用意し加算する．ラベル埋め込みを加算したベースエンコーダから得られる特徴量ベクトル $\mathbf{g}_n^0$ と，近傍事例の特徴量ベクトル $\mathbf{h}_n^k, k \in \{1, 2, \cdots, K\}$ を連結し，行列 $H_n = [\mathbf{g}_n^0, \mathbf{h}_n^1, \cdots, \mathbf{h}_n^K]^T \in \mathbb{R}^{(K+1)\times d'}$ とする．d' は $\mathbf{h}_n^k$ の特徴量次元を示す．なお，本手法はベースエンコーダから得られる特徴量ベクトル $\mathbf{g}_n^0$ と，近傍事例の特徴量ベクトル $\mathbf{h}_n^k$ を連結するため，両者の次元数は同一である必要がある．

次に，自己注意層と FFN 層を用いて，スカラー値 $\lambda_n^k, k \in \{0, 1, \cdots, K\}$ を取得する．行列 H_n を入力し，自己注意層を用いて，式 (5.4) のとおり特徴量ベクトル間の関連性 H_n' を計算する．

$$H_n' = \mathrm{softmax}\left(\frac{H_n W^Q \left(H_n W^K\right)^T}{\sqrt{d'}}\right) H_n W^V \tag{5.4}$$

W^Q, W^K, W^V は自己注意層のパラメータ，$\frac{1}{\sqrt{d'}}$ はスケーリングのパラメータである．また，近傍事例は距離の近い事例から順に取得する．そのため，距離の遠い事例から近い事例へ，影響が及ぶことを防ぐためのマスクを自己注意層に適用する．このマスクは，k $(0 \leq k \leq K)$ 番目の特徴量は，k 以下の j $(j \in \{0, 1, \cdots, k\})$ 番目の特徴量にのみ依存することを示す．

得られた行列 $H_n' = [\mathbf{g}_n^{0\prime}, \mathbf{h}_n^{1\prime}, \cdots, \mathbf{h}_n^{K\prime}]^T$ の内，k 番目の特徴量ベクトルを，ReLU 活性化関数を間に挿入した 2 層の FFN に入力し，式 (5.5) によってスカラー値 λ_n^k を取得する．

$$\lambda_n^k = \begin{cases} \mathrm{ReLU}(\mathbf{g}_n^{0\prime} W_1 + \mathbf{b}_1) W_2 + b_2 & k = 0 \text{ のとき} \\ \mathrm{ReLU}(\mathbf{h}_n^{k\prime} W_1 + \mathbf{b}_1) W_2 + b_2 & 0 < k \leq K \text{ のとき} \end{cases} \tag{5.5}$$

$W_1, W_2, \mathbf{b}_1, b_2$ は FFN のパラメータである．スカラー値 $\lambda_n^k, k \in \{0, 1, \cdots, K\}$ の内，ベースエンコーダの確率分布 p^0 と近傍事例の確率分布 p^K に対応する λ_n^0, λ_n^K を取り出し，Softmax 関数により正規化して重み係数を得る．最後に，正規化した重み係数 λ_n^0, λ_n^K を用いて線形結合を行い，ベースエンコーダによる確率分布 p^0 と近傍事例による確率分布 p^K を式 (5.6) によって組み合わせる．

$$p(y_n|x_{1:n}) = \lambda_n^0 \; p^0(y_n|x_{1:n}) + \lambda_n^K \; p^K(y_n|x_{1:n}) \tag{5.6}$$

動的に重み係数を変更するため，本手法を提案手法 (動的な重み係数) とする．本手法も式 (5.3) に示す提案手法 (静的な重み係数) と同様に，式 (5.6) で得られた確率分布の中で最も確率の高い感情ラベルを，識別結果として出力する．

重み係数の学習　重み係数を導出するネットワークを学習する方法について述べる．5.4.1 項と 5.4.2 項 (ステップ 1) で学習済みのベースエンコーダとクエリーエンコーダのパラメータを固定し，重み係数を導出するネットワークを学習する (ステップ 3)．損失関数として，式 (5.6) の確率分布 p が出力する感情ラベルと教師ラベルとの交差エントロピー (CE: Cross Entropy) 損失を式 (5.7) で計算する．

$$\mathcal{L}_{CE}(\theta) = -\frac{1}{\sum_{m=1}^{M} N_m} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \log \, p\,[y_n] \tag{5.7}$$

ただし，M は訓練データにおける対話の総数を示し，N_m は m 番目の対話の発話数を示す．p は式 (5.6) で組み合わせた確率分布 $p(y_n|x_{1:n})$ を示し，y_n は n 番目の発話の教師ラベルを示す．θ は重み係数を導出するネットワークのパラメータを示す．

本手法はさらに，動的な重み係数を導出するニューラルネットワークを学習するために，発話間の関係を利用するモデルによる確率分布 p^0 と近傍事例による確率分布 p^K のそれぞれが出力する感情ラベルが，教師ラベルと一致する場合に重み係数を高め，そうでない場合に重み係数を低くする損失関数 (係数損失) を導入する．概要を図 5.5 に示す．ベースエンコーダの確率分布 p^0 と近傍事例の確率分布 p^K のそれぞれが出力する感情ラベルが，教師ラベルと一致する場合に 1，そうでない場合に 0 の信号を与え，バイナリー交差エントロピー (BCE: Binary Cross Entropy) 損失を計算する．式 (5.5) に示す Softmax 関数による正規化を行う前のスカラー値 λ_n^k と教師信号を用いて，式 (5.8) に示すシグモイド (Sigmoid) 層とバイナリー交差エントロピー損失を組み合わせた損失関数を計算する．

$$\mathcal{L}_{BCE}(\theta) = -\frac{1}{\sum_{m=1}^{M} N_m} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \mathcal{L}_n \tag{5.8}$$

$$\mathcal{L}_n = \sum_{k \in \{0,K\}} \mathbb{1}_{y_n = \hat{y}_n^k} \log \, \sigma\big(\lambda_n^k\big) + \Big(1 - \mathbb{1}_{y_n = \hat{y}_n^k}\Big) \log\Big(1 - \sigma\big(\lambda_n^k\big)\Big) \tag{5.9}$$

$$\hat{y}_n^k = \arg\max_{c \in \mathbb{Y}} \big(p^k\,[c]\big) \tag{5.10}$$

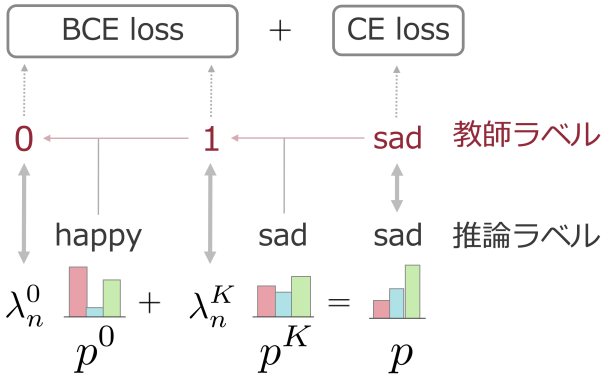

図 5.5: 重み係数の学習方法．最終的な確率分布 p が出力する感情ラベルと，教師ラベルとの交差エントロピー (CE) 損失を計算する．ベースエンコーダによる確率分布 p^0 と近傍事例による確率分布 p^K のそれぞれが，教師ラベルと同じラベルを示す場合に，重み係数を大きく，そうでない場合に重み係数を小さくするように，バイナリー交差エントロピー (BCE) 損失を用いて損失関数を計算する．重み係数を取得するパラメータは交差エントロピー (CE) 損失とバイナリー交差エントロピー (BCE) 損失のマルチタスクで学習する．

ただし，式 (5.10) における p^k は，式 (5.6) に示す発話間の関係を利用するモデルによる確率分布 p^0 または近傍事例による確率分布 p^K を示す．式 (5.10) は確率分布 p^k が出力する感情ラベルを示し，式 (5.9) の $\mathbb{1}_{y_n=\hat{y}_n^k}$ は，教師ラベル y_n とそれぞれの確率分布が出力するラベル $\hat{y}_n^k$ が，一致する場合に 1 を返す指示関数を示す．σ は Sigmoid 関数を示す．

以上より，教師ラベルとの交差エントロピー (CE) 損失と，動的な重み係数を学習するバイナリー交差エントロピー (BCE) 損失とを足し合わせたマルチタスクで，提案手法のパラメータを学習する．最終的な損失関数を式 (5.11) に示す．

$$\mathcal{L}(\theta) = \mathcal{L}_{CE}(\theta) + \mathcal{L}_{BCE}(\theta) \tag{5.11}$$

5.5　実験設定

5.5.1　従来手法との比較

提案手法の有効性を検証するために，以下に示す従来手法と重み付き F1 値を比較する．

KET [Zhong et al., 2019]　階層的自己注意層を用いて文脈を利用する手法である．また，GAT を用いて，常識的知識に関する外部データベースを利用する．

DialogueRNN [Majumder et al., 2019] CNN [Kim, 2014] を用いて発話の特徴量を取得し，話者の特徴と先行文脈，先行発話の感情の関連性について，それぞれ GRU でモデリングする手法である．

DialogueGCN [Ghosal et al., 2019] CNN を用いて発話の特徴量を取得し，隣接する発話間の相互作用を GRU を用いて取得する手法である．加えて，自己依存と他者依存の取得に RGCN と GAT を利用する．

HiTrans [Li et al., 2020a] 発話の内容を示す特徴量の取得に事前学習済み BERT を用い，大域的な文脈の利用に Transformer を用い，階層的に組み合わせた手法である．

DialogXL [Shen et al., 2020] 過去の発話を保存し共有するネットワークを，XLNet [Yang et al., 2019b] に加えた手法である．また隣接する発話間の関係 (局所的) と，会話全体の発話間の関係 (大域的) と，話し手，聞き手の特徴を，それぞれ自己注意層を用いて取得する．

RGAT+P [石渡 et al., 2021] RoBERTa を用いて発話の特徴量を取得し，自己依存と他者依存の取得に RGCN と GAT を利用する手法である．加えて，発話の距離の情報を GAT に組み込む（4.2.2 項）．

COSMIC [Ghosal et al., 2020] 心理的状態や対話の状況に関連する常識的知識を利用する手法である．

RoBERTa 提案手法は，クエリーエンコーダとして RoBERTa を用いる (5.4.2 項)．図 5.4 に示す確率分布 q から得られる感情ラベルを，RoBERTa 単体の出力として用いる．

DAG-ERC [Shen et al., 2021] 話者自身の離れた発話からの影響と，他者の近い発話からの影響を利用するために，GAT を拡張した有向非巡回グラフニューラルネットワークを用いる手法である．本実験では，提案手法との比較のため再現実験を行う．

DialogueCRN [Hu et al., 2021] 対話の状況に応じた相互作用と自身の発言による自己依存を利用する手法である．状況や話者の特徴を理解するために，LSTM を利用する．本実験では，提案手法との比較のため再現実験を行う．ただし，事前学習済みモデルが公開されている IEMOCAP，MELD データセットのみ検証する．

アンサンブル (静的)　本手法はベースエンコーダとして DAG-ERC と DialogueCRN をクエリーエンコーダとして RoBERTa を利用する．異なるモデルの出力を組み合わせるため，アンサンブルの一種である．そこで，伝統的なアンサンブル手法と比較する．図 5.4 に示すベースエンコーダによって得られる確率分布 p^0 と，クエリーエンコーダによって得られる確率分布 q を，重み付き線形和によって組み合わせる．近傍事例による確率分布ではなく，クエリーエンコーダによる確率分布を利用する点で提案手法と異なる．重み係数はハイパーパラメータである．

アンサンブル (動的)　ベースエンコーダによる確率分布 p^0 とクエリーエンコーダによる確率分布 q を入力し，FFN を用いて確率分布を作成する．教師ラベルと推論ラベルとの交差エントロピー損失を用いて FFN のパラメータを学習する．

5.5.2　評価方法

ERC における 3 つのベンチマークセット[2]を用いて，提案手法の有効性を検証する．Shen らの手法 [Shen et al., 2021] で用いられた評価指標と同じ，重み付き F1 値を全てのデータセットの評価に用いる．また，ノンパラメトリック検定の一つである並べ替え検定を用いて，有意差を検定する．検定対象の統計量には，5 回実験を行って得た重み付き F1 値の平均値の差を用い，有意水準 5% の片側検定を行った．

5.5.3　モデルの学習

ステップ 1 におけるベースエンコーダとクエリーエンコーダの学習の設定を示す．ベースエンコーダとして利用する DAG-ERC は [Shen et al., 2021] で報告されたパラメータ[3]を用いて，対話の感情認識タスクで学習した．ただし，ハイパーパラメータの中で学習率のみ， [Shen et al., 2021] で報告された値と異なる値を用いる．その学習率は $(5e-5, 1e-5, 5e-6)$ の中から，検証データで最も性能が高くなるものを選択した．DialogueCRN も同様に，[Hu et al., 2021] で報告されたパラメータ[4]を用いて学習した．クエリーエンコーダとして利用する RoBERTa は，ドロップアウト (Dropout) を 0.3 に設定し，損失関数に交差エントロピー損失を用いて学習した．学習率は $(5e-5, 1e-5, 5e-6)$ の中から，検証データで最も性能が高くなるものを選択した．

[2]2021 年 11 月時点で https://github.com/shenwzh3/DAG-ERC に公開されたデータセットを使用

[3]2021 年 11 月時点で https://github.com/shenwzh3/DAG-ERC に記載されたパラメータを使用

[4]2021 年 11 月時点で https://github.com/zerohd4869/DialogueCRN に記載されたパラメータを使用

♯	モデル	手法	タイプ	IEMOCAP	MELD	EmoryNLP
0	KET			59.56	58.18	34.39
1	DialogueRNN			62.75	-	-
2	DialogueGCN			64.18	58.10	-
3	HiTrans			64.50	61.94	36.75
4	DialogXL			65.94	62.41	34.73
5	RGAT+P			65.95	63.12	35.58
6	COSMIC			65.28	**65.21**	38.11
7	RoBERTa			64.58 ± 1.28	63.67 ± 0.50	38.27 ± 0.52
8		アンサンブル	静的	64.58 ± 1.28	63.67 ± 0.50	38.27 ± 0.52
9			動的	64.47 ± 1.19	63.70 ± 0.64	38.28 ± 0.40
10		提案手法	静的	64.64 ± 1.19	63.96 ± 0.68	38.24 ± 0.69
11			動的	64.69 ± 1.18	63.82 ± 0.50	38.24 ± 0.76
12	DAG-ERC			66.51 ± 0.22	63.12 ± 0.12	38.07 ± 0.47
13		アンサンブル	静的	66.51 ± 0.22	64.30 ± 0.66 •	38.27 ± 0.52
14			動的	64.94 ± 0.69	63.67 ± 0.18 •	38.60 ± 0.31
15		提案手法	静的	66.51 ± 0.22	64.39 ± 0.67 •	38.27 ± 0.47
16			動的	67.33 ± 0.57 •	63.34 ± 0.17	**38.92**± 0.61 •
17	DialogueCRN			63.57 ± 0.54	64.35 ± 0.45	-
18		アンサンブル	静的	66.78 ± 0.41 •	64.94 ± 0.73	-
19			動的	64.95 ± 1.78	64.47 ± 0.53	-
20		提案手法	静的	66.61 ± 0.50 •	65.06 ± 0.65	-
21			動的	**68.11**± 0.69 •	65.02 ± 0.60	-

表 5.1: 従来手法と提案手法の比較．ボールド体は各データセットで最も性能が高い値を示す．下線は各ベースエンコーダと各データセットにおいて最も性能が高い値を示す．黒丸はベースエンコーダに対して統計的な有意差が示された値を示す．各値は 5 回の実験による 重み付き F1 値の平均値を示す．

続いてステップ 3 に関連する実験設定を示す．アンサンブル (静的) と提案手法 (静的な重み係数) の重み係数 λ は，$(0, 0.25, 0.5, 0.75, 1)$ の中から検証データで最も重み付き F1 値が高くなるものを選択した．提案手法 (動的な重み係数) の学習率は $(5e-5, 1e-5)$ の中から検証データで最も性能が高くなるものを選択した．RoBERTa と提案手法 (動的な重み係数) は RAdam optimizer [Liu et al., 2019a] を用いて学習した．提案手法 (静的と動的な重み係数) の特徴量ベクトルの次元数は DAG-ERC [Shen et al., 2021] で報告された 1024 とし，近傍事例の数 K は [Zheng et al., 2021] を参考に 32 に設定した．T は $(1, 10, 100, 1000)$ の中から検証データで最も性能が高くなるものに選択した．近傍事例の検索は，faiss [Johnson et al., 2019] を用いた．全ての実験は 5 回行い，実験結果にはその平均値を用いた．

5.6　結果と考察

5.6.1　従来手法との比較

従来手法との比較結果を表 5.1 に示す．従来手法の KET，DialogueRNN，DialogueGCN，HiTrans，DialogXL，RGAT+P，COSMIC の重み付き F1 値は各文献から引用する．また，RoBERTa をベースエンコーダとクエリーエンコーダに用いたアンサンブル (静的と動的) と提案手法 (静的と動的) の結果も示す．提案手法は，図 5.4 に示す DAG-ERC と DialogueCRN が出力する n 番目の発話の確率分布をベースエンコーダの確率分布として利用した．クエリーエンコーダの RoBERTa も対象の n 番目の発話の確率分布を出力するため，その分布をベースエンコーダの確率分布として利用することが可能である．本実験は，RoBERTa をベースエンコーダとクエリーエンコーダに用いる場合の結果と比較し，ベースエンコーダ (DAG-ERC，DialogueCRN) とクエリーエンコーダ (RoBERTa) に異なるモデルを利用する提案手法の有効性も確認する．表 5.1 のボールド体は各データセットで最も性能が高い値を示し，下線は各ベースエンコーダと各データセットにおいて最も性能が高い値を示す．黒丸は DAG-ERC や DialogueCRN のベースエンコーダに対する統計的有意差を示す．

表 5.1 より，IEMOCAP データセットでは重み付き F1 値 68.11 (♯21) を示し，各文献から引用した従来手法だけでなく，RoBERTa，DAG-ERC や DialogueCRN のベースエンコーダ，アンサンブル手法を大きく上回る最高水準の認識精度を示した．また，EmoryNLP においても，重み付き F1 値 38.92 (♯16) となり，最高水準の認識精度を示した．MELD データセットでは，DAG-ERC や DialogueCRN のベースエンコーダを上回り，提案手法の有効性を確認した．以上の結果より，複数のベンチマークデータセットで高い認識性能を有することから，データのサイズや対話に登場する話者の数が異なる場合でも精度良く認識することを確認した．さらに，IEMOCAP と EmoryNLP データセットにおいて，提案手法 (動的な重み係数) が DAG-ERC や DialogueCRN のベースエンコーダに対して，統計的に有意な差を示すことを確認した (♯16,♯21)．

また，RoBERTa をベースエンコーダに用いたアンサンブル手法 (♯8,♯9) と提案手法 (♯10,♯11) に比べて，複数のベンチマークデータセットで DAG-ERC と DialogueCRN をベースエンコーダに用いたアンサンブル手法 (♯13,♯14,♯18,♯19) と提案手法 (♯15,♯16,♯20,♯21) の認識精度が高いことから，ベースエンコーダと異なるモデル構造を組み合わせる手法の有効性を確認した．

5.6.2　係数損失の効果

5.4.7 項で導入した係数損失の有効性を分析する．提案手法 (動的な重み係数) において，係数損失を使わない場合と，係数損失を使う場合の結果を表 5.2 に示す．

モデル	係数損失	IEMOCAP		MELD		EmoryNLP
		DialogueCRN	DAG-ERC	DialogueCRN	DAG-ERC	DAG-ERC
ベース	-	63.57 ± 0.54	66.51 ± 0.22	64.35 ± 0.45	63.12 ± 0.12	38.07 ± 0.47
提案手法	-	67.31 ± 0.46	66.49 ± 0.24	64.35 ± 0.44	63.13 ± 0.10	38.31 ± 0.32
	✓	**68.11** ± 0.69	**67.33** ± 0.57	**65.02** ± 0.60	**63.34** ± 0.17	**38.92** ± 0.61

表 5.2: 係数損失の効果．提案手法 (動的な重み係数) において，係数損失を使わない場合，係数損失を使う場合の結果を比較する．ボールド体は最も性能が高い値を示す．各値は 5 回の実験による 重み付き F1 値の平均値を示す．

表 5.2 の結果から，ほとんど全てのデータセットで，係数損失を使わない場合よりも係数損失を使う場合は，約 0.6 ∼ 0.8 程度重み付き F1 値が向上した．これは，係数損失の導入によって誤認識に繋がる 0 や 1 といった極端な値を示す重み係数が減少し，両方の確率分布を利用する 0.5 付近に分布したことが要因である．詳細は 5.6.3 項で議論する．

一方で，MELD データセットで DAG-ERC をベースエンコーダに利用する場合は，重み付き F1 値が約 0.2 の向上となり，他のデータセットとベースモデルに比べて性能改善が限定的であった．これは，係数損失を適用した後も，一部のデータで重み係数が誤認識に繋がる 0 や 1 といった極端な値に分布したことが原因である．その理由を 5.6.3 項で議論する．

5.6.3 重み係数の分析

係数損失の有効性をさらに詳細に分析するために，動的に変更した重み係数の頻度分布を分析する．図 5.6 は，IEMOCAP，MELD，EmoryNLP の 3 つの検証データセットにおける，ベースエンコーダの重み係数 λ_n^0 の頻度分布を示す．DAG-ERC をベースエンコーダとして用い，5.6.2 項の実験で用いた係数損失を使わない場合と，係数損失を使う場合を比較する．さらに，検証セットに付与された教師ラベルとの一致を確認し，正答した場合 (青色) と誤答した場合 (赤色) に分けて頻度分布を示す．

5.6.2 項の表 5.2 の係数損失の効果の結果より，係数損失を用いることで全てのデータセットの認識性能が向上した．図 5.6 の結果から，全てのデータセットで，1 付近に分布していた重み係数が 0.5 付近に変化したことがわかる．先行研究 [Kaneko et al., 2022] が示すように，0 や 1 付近に分布する重み係数すなわち片方の分布への依存は性能の劣化を示す傾向にあるため，0.5 付近に分布する重み係数すなわち両方の確率分布を採用する方向に変化したことで，認識性能が向上したことがわかる．

次に，MELD データセットにおける結果を比較する．MELD データセットは，5.6.2 項の表 5.2 の係数損失の効果の結果より，性能の改善が限定的であった．図 5.6 の結果から，係数損失によって重み係数が 0.5 付近に変化した．しかし，1 周辺に

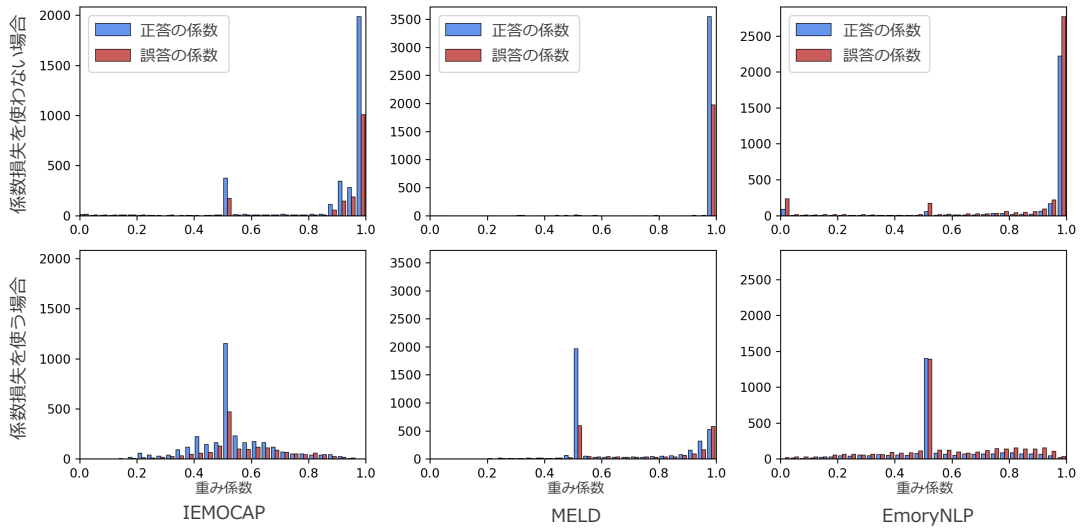

図 5.6: 重み係数の分布．IEMOCAP， MELD， EmoryNLP の 3 つの検証データセットにおける，ベースエンコーダ側の重み係数 λ_n^0 の頻度分布を示す．係数損失を使わない場合，係数損失を使う場合の結果を比較する．教師ラベルと推論ラベルを比較し，正答した場合 (青色) と誤答した場合 (赤色) の係数も比較する．

赤色で示される誤答の係数が存在する．これは，近傍事例の距離が，重み係数の導出に悪影響を与えたことが原因である．

MELD データセットにおいて，1 周辺に赤色で示される誤答の係数が分布する原因をさらに詳細に分析するために，重み係数と近傍事例の距離の等高線図を示す．図 5.7 は，3 つのベンチマークの検証セットにおける，ベースエンコーダ（DAG-ERC）側の重み係数と，5.4.5 項に示す近傍検索によって得られた K 個の事例の距離の平均値の分布を示す．さらに，検証セットに付与された教師ラベルとの一致を確認し，正答した場合 (青色) と誤答した場合 (赤色) に分けて示す．

図 5.7 の結果から，いずれのデータセットも近傍事例の距離が遠い場合に，1 周辺の極端な値に分布し赤色の誤った識別結果に繋がる重み係数が増加することが分かる．すなわち，近傍事例の距離の遠さが，しばしば重み係数の導出に悪影響を与えることが分かる．具体的な事例分析を，5.6.4 項で議論する．

以上の結果をまとめると，重み係数を学習する際は，0.5 付近に重み係数を分布させる係数損失が有効である．しかし，特定のデータセットやベースエンコーダに対する係数損失の有効性は限定的である． これは，近傍事例の距離が遠い場合に，悪影響を及ぼすことが原因である．今後の展望として，重み係数が 0 や 1 といった極端な値を示すことを防ぐための制約を導入した，重み係数導出ネットワークの学習方法を検討する．ネットワークを学習した結果，極端な値を示す重み係数が減少し，より適切な係数を推定することができれば，更なる認識性能の向上が期待できる．

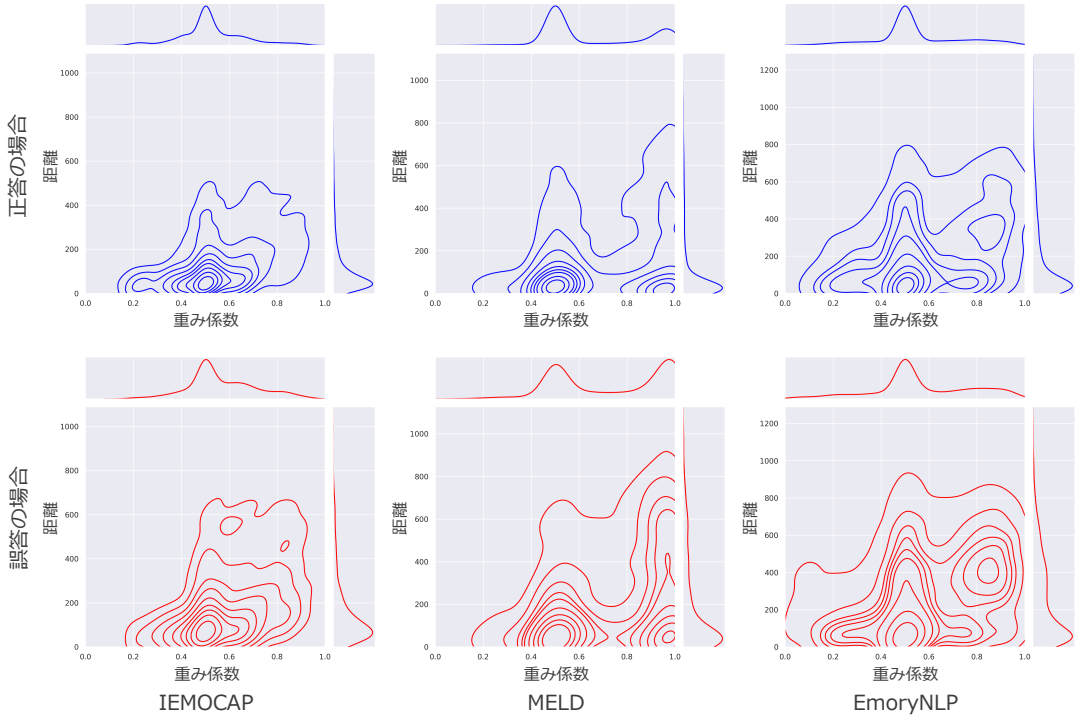

図 5.7: 重み係数と近傍事例の距離の等高線図．ベースエンコーダ（DAG-ERC）側の重み係数と，K 個の近傍事例の距離の平均値の分布を示す．教師ラベルと推論ラベルを比較し，正答した場合 (青色) と誤答した場合 (赤色) を比較する．

5.6.4 事例分析

続いて，ベースエンコーダとクエリーエンコーダの確率分布を組み合わせるアンサンブルと，ベースエンコーダと近傍事例を組み合わせる提案手法 (静的な重み係数)，さらに重み係数を動的に変更する提案手法 (動的な重み係数)，それぞれの特徴を分析するために，事例分析を行う．分析に使用したデータは，MELD データセットの検証セットの一部で，ベースエンコーダによる確率分布と近傍事例による確率分布が示す感情ラベルが異なり，提案手法 (動的な重み係数) が正しく識別した例を示す．図 5.8 は，“*neutral*” が付与された発話に対する各手法の確率分布を示す．1 列目は各手法のベースエンコーダ (DialogueCRN) による確率分布を示す． 横軸は感情ラベルを，縦軸は確率値を示す．“neu” は *neutral*， “hap” は *happy*，“sur” は *surprise*， “sad” は *sadness*，“ang” は *anger*， “dis” は *disgust*，“fea” は *fear* を示す．2 列目は，提案手法 (静的と動的) における近傍検索した各事例と検索クエリーとの距離を示す．横軸は近傍事例の Index を，縦軸はクエリーとの距離を示し，色は感情ラベルの種類を示す．3 列目はクエリーエンコーダまたは近傍事例による確率分布を，4 列目に各手法の最終的な確率分布を示す．各図のタイトルに，各確率分布の重み係数を示す． 表 5.3 に，図 5.8 の分析に使用した識別対象の発話と先行文脈，加えて近傍検索した上位 3 つの事例の発話とラベル，

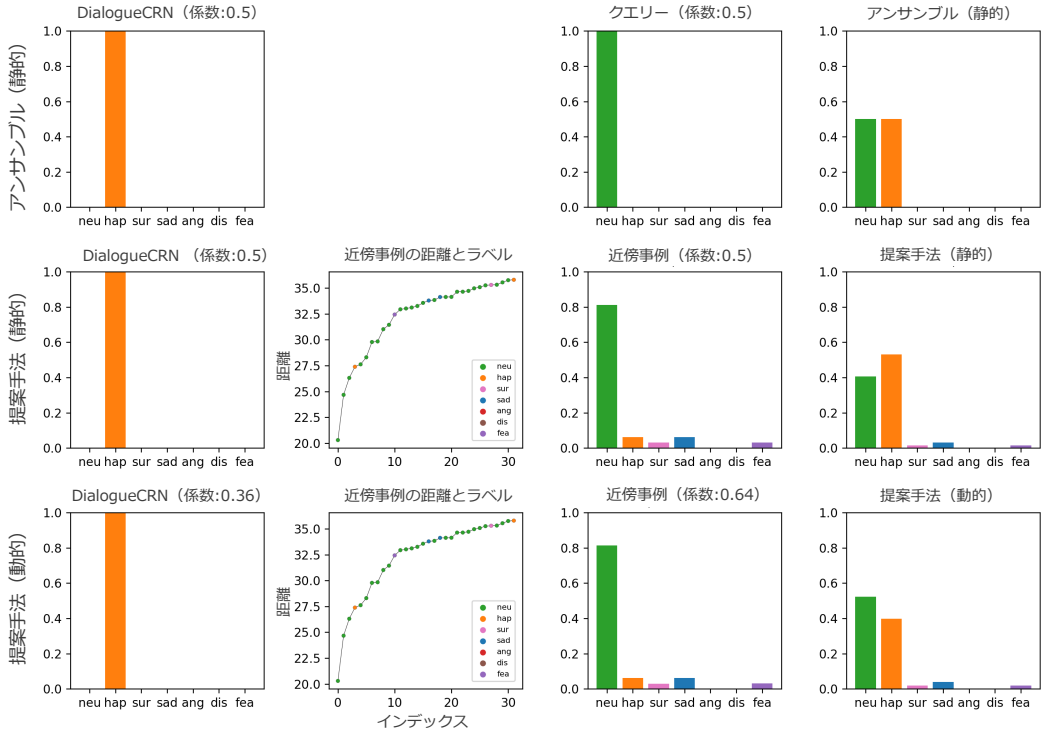

図 5.8: アンサンブルと提案手法 (静的と動的な重み係数) の推定結果の分析．1 列目は各手法のベースエンコーダ (DialogueCRN) による確率分布を示す．2 列目は K 個の近傍事例を示す．3 列目は各手法のクエリーエンコーダまたは近傍事例による確率分布を，4 列目は各手法の最終的な確率分布を示す．各図のタイトルに，各確率分布の重み係数を示す．図は MELD データの検証セットの一部で，*neutral* が付与されたデータである．

検索クエリーとの距離を示す．ボールド体は，識別対象の発話と，それぞれの近傍事例の対象発話を示す．

図 5.8 の結果が示すように，DialogueCRN を用いたベースエンコーダによる確率分布は “*happy*” を示すが，Fintuned RoBERTa を用いたクエリーエンコーダによる分布と近傍事例による分布は，ベースエンコーダと異なるモデルを利用したため異なる感情ラベルを示す場合がある．また，アンサンブルと提案手法 (静的な重み係数) は常に一定の重み係数 0.5 で確率分布を組み合わせるため，ベースエンコーダとクエリーエンコーダ (近傍事例) の確率分布を比較し，高い確率の感情ラベルを最終的に採用する傾向にある．そのため，アンサンブルと提案手法 (静的な重み係数) は最も確率値の高い “*happy*” を出力し，誤認識してしまった．

一方で，提案手法 (動的な重み係数) は，適切な重み係数を算出したため，正しく識別することができた． 表 5.3 の結果が示すように，提案手法 (動的な重み係数) は，識別対象の発話と内容が近い，すなわち距離が近い事例を検索した．距離が近い事例を検索し，結果的に図 5.8 に示すように近傍事例側の重み係数を 0.64 に増加させたことで，この例を正しく識別した．

先行文脈と識別対象の発話	正解ラベル
“Here we go. Okay, brace yourselves.”, “**What?**”	*neutral*

近傍事例		
先行文脈と対象発話	ラベル	距離
“··· Okay! How would you like some Tiki Death Punch?”, “**What's that?**”	*neutral*	20.30
“**What?**”	*neutral*	24.67
“Oh! Hey, Mr. Treeger.”, “**: What are you doing?**”	*neutral*	26.31
⋮		

表 5.3: 図 5.8 の分析に使用した識別対象の発話と先行文脈，近傍検索した上位 3 つの事例の発話とラベル，検索クエリーとの距離．ボールド体は対象の発話を示す．

先行文脈と識別対象の発話	正解ラベル
···, “Yeah, sweetie.”, “**I mean we're not, we're not gonna live together anymore?**”	*sad*

近傍事例		
先行文脈と対象発話	ラベル	距離
···,“Yeah.”, “Hey!”, “**I tried to reach you at work. There's....been a fire.**”	*sad*	122.2
···,“I didn't know there were docks.”, “Hey.”, “Hey.”,“**Aww, is it broken?**”	*sad*	142.5
···,“I want him to have his uncle.”, “**Is my baby gonna have his Uncle Joey?**”	*sad*	154.4
⋮		

表 5.4: 図 5.9 の分析に使用した識別対象の発話と先行文脈，近傍検索した上位 3 つの事例の発話とラベル，検索クエリーとの距離．

次に，提案手法 (動的な重み係数) が誤って識別した例を示す．図 5.9 は，MELD データセットの一部で，“*sad*” が付与された発話に対する各手法の確率分布を示す．ベースエンコーダは，5.6.2 項，5.6.3 項の実験で提案手法 (動的な重み係数) の性能改善が限定的であった DAG-ERC を利用する． 表 5.4 に，図 5.9 の分析に使用した識別対象の発話と先行文脈，加えて近傍検索した上位 3 つの事例の発話とラベル，検索クエリーとの距離を示す．

図 5.9 の結果が示すように，アンサンブルは重み係数 0.5 を，提案手法 (静的な重み係数) は重み係数 0.25 を選択した．アンサンブルと提案手法（静的な重み係数）の両手法は，両方の確率分布を利用することで，この事例を正しく識別した．一方で，提案手法 (動的な重み係数) は，重み係数 0.93 を導出し， ベースエンコーダによる確率分布に極端に依存したため，誤って識別した．表 5.4 の結果が示すように，提案手法 (動的な重み係数) は，識別対象の発話と表層的な内容が類似しない，すなわち距離が遠い事例を検索した．距離が遠い近傍事例を検索し，結果的に図 5.9 に示すように近傍事例側の重み係数を 0.07 に減少させたことで，この例を誤って識別した．

この例を正しく識別するためには，近傍事例の距離の影響を考慮する必要がある．本手法は対話の感情認識タスクで学習したクエリーエンコーダ（5.4.2 項）を

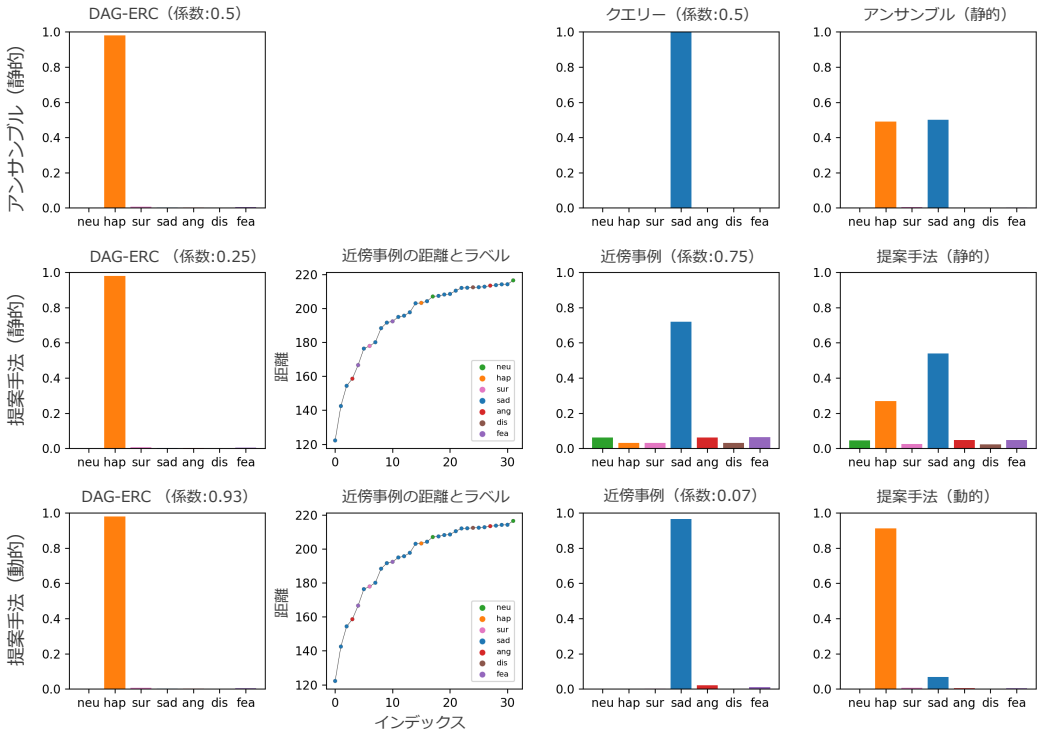

図 5.9: アンサンブルと提案手法 (静的な重み係数)，提案手法 (動的な重み係数) の推定結果の分析．図は MELD データの検証セットの一部で，*sad* が付与されたデータである．

利用するため，表 5.4 の近傍事例が示すように，検索した事例の距離が遠い（発話の内容が類似しない）場合でも，正解ラベルと同じラベルが付与された事例を検索することができる．従って，近傍事例の距離の影響を考慮し，重み係数が極端な値を示すことを防ぐための新たな学習方法を導入することで，より一層の認識性能の向上が期待できる．

5.6.5　クエリーエンコーダーの比較

本手法はクエリーエンコーダとして，Finetuned RoBERTa を利用した．Finetuned RoBERTa の有効性を確認するために，ベースエンコーダ (DAG-ERC，DialogueCRN) と同じモデルをクエリーエンコーダに用いる場合，再学習をしない RoBERTa(Vanilla) を用いる場合の結果を比較する．提案手法は図 5.4 に示す RoBERTa が出力する特徴量ベクトルを検索クエリーに利用した．ベースエンコーダも対象の n 番目の発話の特徴量ベクトルを出力するため，その特徴量を検索クエリーに利用することが可能である．本実験は，ベースエンコーダと同じモデルをクエリーエンコーダに用いる手法と比較し，Finetuned RoBERTa をクエリーエンコーダに用いる提案手法の有効性を確認する．また，Finetuned RoBERTa をベー

♯	ベース	タイプ	クエリー	IEMOCAP	MELD	EmoryNLP
0	Finetuned			64.58 ± 1.28	63.67 ± 0.50	38.27 ± 0.52
1		静的	Finetuned	64.64 ± 1.19	63.96 ± 0.68	38.24 ± 0.69
2		動的	Finetuned	64.69 ± 1.18	63.82 ± 0.50	38.24 ± 0.76
3	DAG-ERC			66.51 ± 0.22	63.12 ± 0.12	38.07 ± 0.47
4			DAG-ERC	66.51 ± 0.22	63.12 ± 0.12	38.06 ± 0.48
5		静的	Vanilla	66.51 ± 0.22	63.12 ± 0.12	38.07 ± 0.47
6			Finetuned	66.51 ± 0.22	**64.39** ± 0.67	38.27 ± 0.47
7			DAG-ERC	66.59 ± 0.20	63.14 ± 0.08	38.19 ± 0.39
8		動的	Vanilla	66.48 ± 0.23	63.13 ± 0.12	38.02 ± 0.50
9			Finetuned	**67.33** ± 0.57	63.34 ± 0.17	**38.92** ± 0.61
10	DialogueCRN			63.57 ± 0.54	64.35 ± 0.45	-
11			DialogueCRN	65.27 ± 0.39	64.36 ± 0.32	-
12		静的	Vanilla	63.86 ± 0.42	64.51 ± 0.31	-
13			Finetuned	66.61 ± 0.50	**65.06** ± 0.65	-
14			DialogueCRN	65.00 ± 0.35	64.28 ± 0.45	-
15		動的	Vanilla	63.66 ± 0.42	64.46 ± 0.46	-
16			Finetuned	**68.11** ± 0.69	65.02 ± 0.60	-

表 5.5: クエリーエンコーダの比較．ベースエンコーダと同じ手法，再学習をしない RoBERTa (Vanilla)，Finetuned RoBERTa (Finetuned) の比較．ボールド体は各データセットと各ベースエンコーダで最も性能が高い値を示す．

スエンコーダとクエリーエンコーダの両方に利用する手法も比較する．本実験は提案手法 (静的) と提案手法 (動的) の両手法で検証する．実験結果を表 5.5 に示す．

表 5.5 より，提案手法 (静的な重み係数) と提案手法 (動的な重み係数) の両方で，クエリーエンコーダとして Finetuned RoBERTa を用いる手法は，最も高い認識性能を示した．ベースエンコーダと同じモデルをクエリーエンコーダに用いた手法 (♯4,♯7,♯11,♯14) と Finetuned RoBERTa をベースエンコーダとクエリーエンコーダに用いた手法 (♯1,♯2) と比較して，Finetuned RoBERTa をクエリーエンコーダに用いた手法 (♯6,♯9,♯13,♯16) の性能が高いことから，ベースエンコーダと異なるモデルを組み合わせる手法の有効性を確認できる．一方で，ベースエンコーダと同じモデルをクエリーエンコーダに用いる手法 (♯4,♯7,♯11,♯14) は，ベースエンコーダ単体 (♯3,♯10) に対して，Finetuned RoBERTa をクエリーエンコーダに用いた手法 (♯6,♯9,♯13,♯16) に相当する性能の改善は認められない．これは，ベースエンコーダと同じモデルをクエリーエンコーダに利用したことで，ベースエンコーダによる確率分布 p^0 と近傍事例による確率分布 p^K の相関が高くなり，アンサンブルの効果が限定的になってしまったためである．また，クエリーエンコーダとして再学習を行わない RoBERTa (Vanilla) を用いた手法 (♯5,♯8,♯12,♯15) と比較し，Finetuned RoBERTa を用いた手法 (♯6,♯9,♯13,♯16) の性能が高いことから，対話の感情認識の

モデル	確率分布の種類 ベース	クエリ	kNN	IEMOCAP	MELD	EmoryNLP
DAG-ERC	✓	-	-	66.51 ± 0.22	63.12 ± 0.12	38.07 ± 0.47
	-	✓	-	64.58 ± 1.28	63.67 ± 0.50	38.27 ± 0.52
	-	-	✓	65.27 ± 1.07	64.13 ± 0.60	37.83 ± 0.74
アンサンブル	✓	✓	-	66.51 ± 0.22	64.30 ± 0.66	38.27 ± 0.52
提案手法 (静的)	✓	-	✓	66.51 ± 0.22	**64.39** ± 0.67	38.27 ± 0.47
提案手法 (動的)	✓	-	✓	**67.33** ± 0.57	63.34 ± 0.17	**38.92** ± 0.61

表 5.6: ベースエンコーダ (ベース)，クエリーエンコーダ (クエリ)，近傍事例 (kNN) による確率分布の，それぞれが出力する感情ラベルの正確さを比較．ベースエンコーダーとして DAG-ERC を用いる．ボールド体は最も性能が高い値を示す．

モデル	確率分布の種類 ベース	クエリ	kNN	IEMOCAP	MELD	EmoryNLP
DialogueCRN	✓	-	-	63.57 ± 0.54	64.35 ± 0.45	-
	-	✓	-	64.58 ± 1.28	63.67 ± 0.50	-
	-	-	✓	65.27 ± 1.07	64.13 ± 0.60	-
アンサンブル	✓	✓	-	66.78 ± 0.41	64.94 ± 0.73	-
提案手法 (静的)	✓	-	✓	66.61 ± 0.50	**65.06** ± 0.65	-
提案手法 (動的)	✓	-	✓	**68.11** ± 0.69	65.02 ± 0.60	-

表 5.7: ベースエンコーダ (ベース)，クエリーエンコーダ (クエリ)，近傍事例 (kNN) による確率分布の，それぞれが出力する感情ラベルの正確さを比較．ベースエンコーダーとして DialogueCRN を用いる．ボールド体は最も性能が高い値を示す．

観点で近い事例を検索することが認識性能の向上に寄与することがわかる．

5.6.6　確率分布の組み合わせによる効果

最後に，ベースエンコーダ (ベース) による確率分布と，クエリーエンコーダ (クエリ) による確率分布と，近傍事例 (kNN) による確率分布の，それぞれが出力する感情ラベルの正確さを分析する．それぞれの確率分布で最も高い感情ラベルを出力と見なし，教師ラベルとの重み付き F1 値を計算した結果を示す．複数の確率分布を組合せた場合と比較するために，ベースエンコーダとクエリーエンコーダを組み合わせるアンサンブルと，ベースエンコーダと近傍事例を組み合わせる提案手法の結果も合せて再掲する．ベースエンコーダーとして DAG-ERC を用いた結果を表 5.6 に，DialogueCRN を用いた結果を表 5.7 に示す．

表 5.7 の結果より，ベースエンコーダとして DialogueCRN を利用する場合，ベースエンコーダ，クエリーエンコーダ，近傍事例による確率分布を単体で利用する

よりも，アンサンブルと提案手法の認識性能が高いことから，異なる確率分布を組み合わせることが精度の向上に寄与することがわかった．

また，表 5.6 と表 5.7 において，IEMOCAP と EmoryNLP データセットにおいて，重み係数を動的に変更する提案手法が最も高い認識性能を示すことを確認した．

5.7 会話の履歴と発話間の関係を組み合わせる手法のまとめ

本章は，対話の感情認識に，近傍事例を活用する手法を初めて適用した．k 近傍法を用いて会話の履歴の観点で意味的に近い発話を訓練セットから検索し，検索した発話 (近傍事例) に付与された感情ラベルを基に確率分布を作成して，発話間の関係を利用するモデルの確率分布と重み付き線形和によって組み合わせた．さらに，定数による重み係数で 2 つの確率分布を足し合わせるだけでなく，識別対象の発話ごとに動的に重み係数を変更する手法を提案した．3 つのベンチマークデータセットによる評価実験を通して，動的に重み係数を変更する提案手法は最高水準の認識性能を示し，有効性を確認した．

第6章　結論

本論文は，対話の感情認識において，ある発話から他の発話の感情に影響を与える発話間の関係と，先行文脈に応じて異なる感情を示す会話の履歴の 2 つの課題に着目し，これらの課題に対処する手法を提案した．対話の感情認識のベンチマークデータセットによる評価実験を通して，提案手法は感情認識の性能向上に貢献することを確認した．

まず，発話間の関係の課題では，発話の距離を利用する手法を提案した．対話の感情認識では，発話間の関係の中で，自分自身の感情の推移を表す自己依存と他者の発話が影響を与える他者依存の関係が感情に影響を与えることが知られている．従来手法の多くは，グラフニューラルネットワークを用いて自己依存と他者依存の関係を利用し，高い認識性能を示した．しかしながら，これらの依存関係を利用する手法は，発話の距離を考慮しないという課題が存在する．話者の感情はしばしば発話から発話への距離に依存する．そこで本論文は，発話間の関係に加えて，対象の発話から周辺の発話への依存関係の種類に応じた距離の情報も利用する手法を提案した．提案手法を用いることで，自己依存と他者依存を含む発話間の関係と，発話の距離の両方を利用できる．対話の感情認識における 3 つのベンチマークデータによる評価実験を通して，提案手法の有効性を確認した．また，依存関係の種類に応じた距離の情報が，対話の感情認識の認識性能の向上に貢献することも確認した．

次に，会話の履歴の課題では，発話間の関係を利用するモデルと会話の履歴を利用するモデルを組み合わせる手法を提案した．対話の感情認識では，同じ発話であっても，一連の会話の履歴に応じて異なる感情を示すことがある．会話の履歴を利用する代表的な方法として，連続した複数の発話を連結し言語モデルに入力する方法がある．この手法は，識別対象の発話とその先行文脈に注意を向けるため，一連の会話の履歴を利用することができる．しかしながら，この手法は，会話全体に注意を向けるため，逆に個々の発話の依存関係の利用が容易でない．そこで，本論文は発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組み合わせるアンサンブル手法を提案した．単純に組み合わせるだけでなく，過去の会話から会話の内容が近いものを検索し，動的な重み付き線形和によって補強する事例ベース手法を提案した．具体的には，識別対象の発話とその先行文脈をクエリーとして，会話の履歴の観点で意味的に近い発話を訓練データセットから k 近傍法を用いて検索した．検索した発話 (近傍事例) に付与された感情ラベルと，識別対象の発話との距離を基に感情ラベルの確率分布を作成し，発話間の関

係を利用するモデルの確率分布と，動的な重み付き線形和によって組み合わせた．提案手法を用いることで，発話間の関係と会話の履歴の両方の特徴を利用することができる．対話の感情認識における 3 つのベンチマークデータによる評価実験を通して，動的に重み係数を変更する提案手法が，最高水準の認識性能を示し，有効性を確認した．

6.1　貢献

本論文の貢献を示す．

6.1.1　発話の距離を利用した発話間の関係

発話の距離を利用する提案手法について，貢献を以下に示す．本論文は，対話の感情認識において，発話と発話の相対的な位置を利用するため，初めて RGAT に距離の情報を付与する方法を提案した．提案手法を用いることで，自己依存と他者依存を含む発話間の依存関係と，発話の距離の両方の利用を可能にした．従来手法との比較実験を通して，提案手法は従来手法を上回る最高水準の認識性能を示し，その有効性を確認した．さらに，評価実験を通して，依存関係の種類に応じた距離の情報の有効性を確認した．

提案手法は，RGAT に距離の情報を付与する方法であり，対話の感情認識タスクだけでなく，知識グラフのノード分類などグラフ構造を利用する様々な研究課題に応用が可能である．

6.1.2　会話の履歴と発話間の関係の組み合わせ

会話の履歴を利用する提案手法について，貢献を以下に示す．本論文は，対話の感情認識において，発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組み合わせる手法を提案した．2 つの異なるモデルを組み合わせる方法として，近傍事例を活用し，初めて対話の感情認識タスクに適用した．単純に組み合わせるだけでなく，識別対象の発話に応じて動的に変化する重み係数を用いて，発話間の関係を利用するモデルの確率分布と，近傍事例による確率分布を組み合わせた．従来手法との比較実験を通して，重み係数を動的に変更する提案手法は，従来手法を上回る最高水準の認識性能を示し，その有効性を確認した．

提案手法は，ベースエンコーダとして発話間の関係を考慮するモデルを利用し，クエリーエンコーダとして会話の履歴を考慮するモデルを利用した．ベースエンコーダやクエリーエンコーダとして，識別対象の発話を表す特徴量ベクトルと感情ラベルの確率分布を出力する他の識別モデルも利用することができる．

6.2 今後の展望

本論文の今後の展望を示し，対話の感情認識タスクに関わる将来の展望を示す．

6.2.1 発話の距離を考慮した発話間の関係

発話の距離を考慮した発話間の関係を利用する手法の今後の展望として，距離の情報を示す変数の次元を増加させることを検討する．本論文は，4.4 節に示すように，依存関係の種類に応じた距離の情報を，RGAT のエッジ重み係数に加算するために，スカラー値で表した．スカラー値では表現次数として不十分な可能性があるので，今後，距離の情報を示す変数の次元を増加させて，より適切な距離の情報を表現させることを検討する．

昨今，発話と発話の依存関係をモデリングする方法として，RGAT に代わる新たな手法 [Shen et al., 2021] が提案されている．しかしながら，この手法は発話と発話の距離を考慮していない．今後は，発話間の依存関係を効果的に利用する新たな手法に，距離の情報を組み合わせる方法を検討する．

6.2.2 会話の履歴と発話間の関係の組み合わせ

会話の履歴と発話間の関係を組み合わせる手法の今後の展望として，重み係数導出ネットワークの学習方法を検討する．本論文は，5.4.7 項に示すように，発話間の関係を利用するモデルと，会話の履歴を利用するモデルを組み合わせるために，動的な重み係数を導入した．前者のモデルが適切な場合はそのモデル側の重み係数を高くする係数損失を導入した．しかしながら，係数損失を利用しても，重み係数が極端な値を示す場合があり，特定のデータセットでその効果が限定的であった．そこで今後は，重み係数が 0 や 1 といった極端な値を示すことを防ぐための制約を加えた，重み係数導出ネットワークの学習方法を検討する．ネットワークを学習した結果，極端な重み係数が減少し，適切な係数を導出できれば，より一層の認識性能の向上が期待できる．

本論文は対話の感情認識の研究課題の中で，発話間の関係と会話の履歴に着目し，それぞれの課題を解決する識別モデルをベースエンコーダやクエリーエンコーダに利用した．他の研究課題として，1.2.3 項に示す話者固有の特徴や常識的知識があり，これらを利用する識別モデルも高い認識性能を示してきた．今後，他の研究課題に用いられる識別モデルをベースエンコーダやクエリーエンコーダに利用することで，認識性能の向上が期待できる．

6.2.3　対話の感情認識

マルチモーダル

マルチモーダルを利用した対話の感情認識について，将来の展望を示す．本論文は，話者の音声を書き起こしたテキストの情報を利用し，対話における各発話の感情を認識するモデルを構築した．しかし，テキストの情報だけでは，話者が表現する複雑な感情を十分に理解できない場合がある．例えば，ある話者は，しばしば怒りの感情を表現する際に，直接言葉で表さずに中立的な発言をする．テキスト情報のみを利用するモデルは，このような発言が示す本来の感情を理解することは容易でない．しかし，話者の表情や声のトーンといった情報を利用できれば，このような複雑な感情を理解できる可能性がある．今後，テキストの情報だけでなく，話者の表情などが現れる映像や，声のトーンなどが現れる音声の情報も利用するマルチモーダルな識別モデルを構築することで，更なる認識性能の向上が期待できる．

大規模言語モデル

大規模言語モデルを利用した対話の感情認識について，将来の展望を示す．テキスト生成の枠組みにより様々な自然言語理解タスクを統合して学習可能となった大規模言語モデルは，あらゆるタスクで大幅な性能の向上を示している．対話コーパスで事前に学習した大規模言語モデルは，発話間の関係など対話の感情認識タスク固有の課題を解決する能力を，既に内包していると考えられる．大規模言語モデルからそれらの能力を効果的に引き出すためには，品質の高いプロンプトの作成が求められる．今後，対話の文脈や話者固有の情報などの対話の感情認識タスク固有の情報を効果的に引き出すプロンプトを設計し，大規模言語モデルに利用することで，更なる認識性能の向上が期待できる．

参考文献

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*, pages 2206–2240. PMLR, 2022.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335, 2008.

Sheng-Yeh Chen, Chao-Chun Hsu, Chuan-Chun Kuo, Lun-Wei Ku, et al. Emotionlines: An emotion corpus of multi-party conversations. *arXiv preprint arXiv:1802.08379*, 2018.

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018. https://arxiv.org/abs/1810.04805.

Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4):384, 1993.

Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation. 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 154–164, 2019. URL https://www.aclweb.org/anthology/D19-1015/.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. Cosmic: Common sense knowledge for emotion identification in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, 2020. doi: 10.18653/v1/2020.findings-emnlp.224. URL https://aclanthology.org/2020.findings-emnlp.224.

Qiuxiang He, Guoping Huang, Qu Cui, Li Li, and Lemao Liu. Fast and accurate neural machine translation with translation memory. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3170–3180, 2021.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7042–7052, 2021. doi: 10.18653/v1/2021.acl-long.547. URL https://aclanthology.org/2021.acl-long.547.

John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. In *Advances in Neural Information Processing Systems*, pages 15794–15805, 2019. URL https://papers.nips.cc/paper/9711-generative-models-for-graph-based-protein-design.pdf.

Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 7360–7370, 2020.

Qingnan Jiang, Mingxuan Wang, Jun Cao, Shanbo Cheng, Shujian Huang, and Lei Li. Learning kernel-smoothed machine translation with retrieved examples. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7280–7290, 2021. URL https://aclanthology.org/2021.emnlp-main.579/.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R. Lyu. Higru: Hierarchical gated recurrent units for utterance-level emotion recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 397–406, Minneapolis, Minnesota, 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1037. URL https://aclanthology.org/N19-1037.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.

Masahiro Kaneko, Sho Takase, Ayana Niwa, and Naoaki Okazaki. Interpretability for language learners using example-based grammatical error correction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7176–7187, Dublin, Ireland, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.496. URL https://aclanthology.org/2022.acl-long.496.

Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

Urvashi Khandelwal, Angela Fan, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Nearest neighbor machine translation. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=7wCBOfJ8hJM.

Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1181. URL https://aclanthology.org/D14-1181.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*, 2016.

Joosung Lee and Wooin Lee. Compm: Context modeling with speaker's pre-trained memory tracking for emotion recognition in conversation. In *Proceedings of the 2022 Conference of the*

North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5669–5679. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.naacl-main.416. URL https://aclanthology.org/2022.naacl-main.416.

Jingye Li, Donghong Ji, Fei Li, Meishan Zhang, and Yijiang Liu. Hitrans: A transformer-based context-and speaker-sensitive model for emotion detection in conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4190–4200, 2020a.

Qingbiao Li, Chunhua Wu, Zhe Wang, and Kangfeng Zheng. Hierarchical transformer network for utterance-level emotion recognition. *Applied Sciences*, 10(13):4447, 2020b.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In *Proceedings of the International Conference on Learning Representations*, 2019a.

Xiao Liu, Jian Zhang, Heng Zhang, Fuzhao Xue, and Yang You. Hierarchical dialogue understanding with special tokens and turn-level attention. *arXiv preprint arXiv:2305.00262*, 2023.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. 2019b. https://arxiv.org/abs/1907.11692.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *Proceedings of the International Conference on Learning Representations*, 2016.

Xin Lu, Yanyan Zhao, Yang Wu, Yijian Tian, Huipeng Chen, and Bing Qin. An iterative emotion interaction network for emotion recognition in conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4078–4088, 2020.

Linkai Luo and Yue Wang. Emotionx-hsu: Adopting pre-trained bert for emotion classification. 2019. https://arxiv.org/abs/1907.09669.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825, 2019. doi: 10.1609/aaai.v33i01.33016818. URL https://arxiv.org/abs/1811.00405.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. Mime: Mimicking emotions for empathetic response generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8968–8979, 2020. URL https://aclanthology.org/2020.emnlp-main.721/.

Patrícia Pereira, Helena Moniz, and Joao Paulo Carvalho. Deep emotion recognition in textual conversations: A survey. *arXiv preprint arXiv:2211.09172*, 2022.

Robert Plutchik. A psychoevolutionary theory of emotions. *Social Science Information*, 21: 529–553, 1982.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 873–883, 2017. doi: 10.18653/v1/P17-1081. URL https://www.aclweb.org/anthology/P17-1081/.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, 2018.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, 7(1):100943–100953, 2019.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference Springer*, pages 593–607, 2018. doi: 10.1007/978-3-319-93417-4_38. URL https://arxiv.org/abs/1703.06103.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, page 464–468, 2018.

Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition. In *35th AAAI Conference on Artificial Intelligence*, 2020.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. Directed acyclic graph network for conversational emotion recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1551–1560, 2021. doi: 10.18653/v1/2021.acl-long.123. URL https://aclanthology.org/2021.acl-long.123/.

Dongming Sheng, Dong Wang, Ying Shen, Haitao Zheng, and Haozhuang Liu. Summarize before aggregate: a global-to-local heterogeneous graph inference network for conversational emotion recognition. In *Proceedings of the 28th international conference on computational linguistics*, pages 4153–4163, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *Proceedings of the International Conference on Learning Representations*, 2017.

Dexin Wang, Kai Fan, Boxing Chen, and Deyi Xiong. Efficient cluster-based k-nearest-neighbor machine translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2175–2187, 2022a. URL https://aclanthology.org/2022.acl-long.154/.

Shuhe Wang, Xiaoya Li, Yuxian Meng, Tianwei Zhang, Rongbin Ouyang, Jiwei Li, and Guoyin Wang. k nn-ner: Named entity recognition with nearest neighbor search. 2022b. https://arxiv.org/abs/2203.17103.

Xing Wang, Zhaopeng Tu, Longyue Wang, and Shuming Shi. Self-attention with structural position representations. 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), page 1403–1409, 2019. URL https://www.aclweb.org/anthology/D19-1145/.

Kisu Yang, Dongyub Lee, Taesun Whang, Seolhwa Lee, and Heuiseok Lim. Emotionx-ku: Bert-max based contextual emotion classifier. 2019a. https://arxiv.org/abs/1906.11565.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019b.

Sayyed M Zahiri and Jinho D Choi. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. URL https://arxiv.org/abs/1708.04299.

Xin Zheng, Zhirui Zhang, Junliang Guo, Shujian Huang, Boxing Chen, Weihua Luo, and Jiajun Chen. Adaptive nearest neighbor machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 368–374, 2021. URL https://aclanthology.org/2021.acl-short.47/.

Peixiang Zhong, Di Wang, and Chunyan Miao. Knowledge-enriched transformer for emotion detection in textual conversations. 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)*, 2019.

Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*, pages 4623–4629, 2018.

太智 石渡, 有希 安田, 太郎 宮﨑, and 淳 後藤. 発話順序に基づく graph attention networks を用いた 対話文における感情認識. **自然言語処理**, 28(4):1141–1161, 2021. doi: 10.5715/jnlp.28.1141.

太智 石渡, 淳 後藤, 寛章 山田, and 健伸 徳永. 近傍事例を用いた対話における感情認識. **自然言語処理**, 31(2):504–533, 2024. doi: 10.5715/jnlp.31.504.

謝辞

本研究を遂行し博士論文を作成するにあたり，多くの方々からご指導とご支援をいただきました．この場をお借りして厚くお礼申し上げます．

指導教員の徳永健伸教授には，博士課程の 3 年間を通して多くの指導を賜りました．論文の作成に不慣れな私に対して，基本から丁寧にご指導をいただき，論理的かつ明確な文章を書く力を身につけることができました．なかなか研究成果が出ずに悩むことが多かった時期も，徳永先生の根気強いご助言により研究を無事に進めることができました．また，ポイントを明確にして簡潔にプレゼンする能力を養うことができたのは，徳永先生のご指導のおかげです．これらの経験を，今後の研究や仕事において大いに役立てていきたいと考えております．これまでにご指導をいただきまして誠にありがとうございました．

本論文の審査を引き受けてくださった村田剛志教授，宮崎純教授，岡崎直観教授，齋藤豪准教授に心より感謝申し上げます．村田剛志教授には，既存研究との位置付けについて的確なコメントをいただき，本論文の位置付けをより明確にすることができました．また，宮崎純教授には，審査を通して，疑問点や不明確な部分を残さずに原因を突き詰めていく姿勢を学ぶ機会をいただきました．この心構えは今後の研究活動においても大切にしてまいります．岡崎直観教授には，的確かつ核心をつくご指摘をいただき，本研究をさらに深めることができました．齋藤豪准教授には，博士論文をより良くするための有益なコメントをいただき，論文の完成度を高めることができました．

徳永研究室の皆様に，感謝の意を表します．山田寛章助教授には，研究の進め方や，より効果的にプレゼンテーションを行う方法など，多くのご指導をいただきました．特に，仮説を検証するための実験の設計や，視覚的かつ効果的なプレゼンを行う工夫に関するアドバイスは，私にとって大変貴重なものとなりました．心より感謝申し上げます．徳永研究室に所属する皆様には，学会投稿時のピアレビューで多大なご協力をいただき，論文の完成度を高めることができました．また，本研究をより深く掘り下げることができたのは，プレナリーミーティングでの活発な質疑のおかげです．

日本放送協会の皆様には大変お世話になりました．山田一郎さん，後藤淳さん，宮崎太郎さんには，社会人博士を進めるにあたり，研究に専念できる職場環境の整備にご協力をいただきました．皆様のお力添えなくしては，学業と仕事の両立はできませんでした．心から感謝しております．徳永研究室を卒業された美野秀弥さんには， 研究だけでなく社会人博士を進める上で，必要な情報を多くいただ

きました．同じ研究部に所属するみなさまとの，日々の研究に関する議論や雑談は，精神的な支えとなりました．この場をお借りして感謝申し上げます．

最後に，博士課程を進めるに際して，サポートしてくれた家族，特に妻には，仕事と学業の両立で忙しい毎日を過ごす中で，日々の生活を支え研究に集中できる環境を整えてくれたことに感謝しております．