

概要

コールセンターにおいて、ユーザの状態の把握やトラブルの解決に時間がかかることが問題となっている。このことからコールセンターにおいてチャットボットの導入が検討されている。そこで村田ら [1] は、情報抽出の技術を利用し、コールセンター業務支援となるチャットボットへの応用に役立つようなフローチャート作成を行った。しかし作成されるフローチャートには類似箇所が複数出現する場合があった。そこで本研究では、類似箇所のマージを行うことでフローチャートを簡略化する。

村田らの研究では、情報抽出の技術を用いコールセンターでの対話データを表の形に整理し、表から対話に役立つフローチャートの作成を行っていた。

本研究では、フローチャートの類似箇所をマージすることによって、フローチャートの規模を縮小し、フローチャートを簡略化する手法を提案する。マージの判定は類似箇所の対話の内容を基準に3種類の判定方法で行う。マージを行うことで、フローチャートの規模が縮小し、チャットボットの管理を行う際に対話の全体像を素早く把握できるフローチャートになると考えられる。

「電源がつかない」の問診 20 件と「再生中にフリーズする」の問診 20 件で提案手法の3種類の判定方法で実験を行った結果、「電源がつかない」のフローチャートのノード数が97個からそれぞれ25個、64個、56個に減少し、「再生中にフリーズする」のフローチャートのノード数が74個からそれぞれ24個、47個、44個に減少した。3種類のマージの判定方法全てでフローチャートのノード数が減少し、フローチャートの規模が縮小した。

フローチャートの簡略化は行えたが、マージをする際にフローチャートの対話内容に変化が起こることがあった。情報が増加しより良い内容になることもあったが、内容が変化し対話の流れが不自然になる場合もあった。類似箇所をマージしているためこのようなことが生じる。

今後の課題として、類似箇所でもマージを行っても対話が不自然にならないマージ箇所の判定方法を見つけることが考えられる。

目次

第1章	はじめに	1
第2章	関連研究	2
2.1	表作成の手順	2
2.2	フローチャートの自動作成の手順	3
2.2.1	MF 基本手法	3
2.2.2	MF ダミー手法	4
2.2.3	MF 単一回答削除手法	5
2.2.4	MF ダミー削除手法	5
第3章	提案手法	7
3.1	マージを行う候補	7
3.2	マージの判定	9
3.2.1	MF 同一質問ノード融合手法	9
3.2.2	MF 同一後続一致ノード融合手法	11
3.2.3	MF 同一後続類似ノード融合手法	13
第4章	実験	16
4.1	実験データ	16
4.2	評価方法	17
4.3	実験結果	17
第5章	考察	18
5.1	マージの長所	18
5.2	マージの短所	22
第6章	今後の課題	26

表 目 次

2.2.1 対話データから作成された表	3
3.1.1 マージ前の表	7
3.2.1 MF 同一質問ノード融合手法でのマージ例	9
3.2.2 完全一致している箇所を含む表	11
3.2.3 完全一致している箇所をマージした表	12
3.2.4 類似箇所の内容が異なる時の表	13
3.2.5 MF 同一後続類似ノード融合手法でマージした表	14
4.1.1 問診の内容 (一部)	16
4.3.1 ノード数の変化	17

目 次

2.1	表から作成されたフローチャートの例	4
2.2	ダミーノードを含む対話	4
2.3	MF 単一回答削除手法	5
2.4	MF ダミー削除手法	6
3.1	マージ前の表から作成されるフローチャート	8
3.2	MF 同一質問ノード融合手法でマージしたフローチャート	10
3.3	完全一致している箇所を含むフローチャート	11
3.4	完全一致している箇所をマージしたフローチャート	12
3.5	類似箇所の内容が異なる時のフローチャート	14
3.6	MF 同一後続類似ノード融合手法でマージしたフローチャート	15
5.1	マージ前の対話	18
5.2	マージ後の情報が追加された対話	19
5.3	完全一致している箇所を含む対話	19
5.4	完全一致している箇所をマージした対話	20
5.5	完全一致ではマージしなかった対話	20
5.6	MF 同一後続類似ノード融合手法でマージした対話	21
5.7	マージ前のユニークな対話	22
5.8	ユニークな対話をマージした対話	23
5.9	MF 同一後続一致ノード融合手法ではマージできない対話	24
5.10	異なる内容の質問が類似質問となされている対話	25
5.11	異なる内容である質問がマージされた対話	25

第1章 はじめに

コールセンターにおいて、ユーザの状態の把握やトラブルの解決に時間がかかることが問題となっている。このことからコールセンターにおいてチャットボットの導入が検討されている。そこで村田ら [1] は、情報抽出の技術を利用し、コールセンター業務支援となるチャットボットへの応用に役立つようなフローチャートの作成を行った。しかし作成されるフローチャートには類似箇所が複数出現することで、フローチャートの規模が大きくなり、フローチャートが見にくくなる場合があった。そこで本研究では、類似箇所のマージを行うことでフローチャートを簡略化する。

村田ら [1] の研究では、情報抽出の技術を用いコールセンターでの対話データを表の形に整理し、表から対話に役立つフローチャートの作成を行っていた。

本研究では対話データを表の形に整理した後、その表で類似箇所のマージを行うことで、フローチャートを簡略化する手法を提案する。マージを行うことで、フローチャートの規模が縮小し、チャットボットの管理などを行う際に対話の全体像を素早く把握できるフローチャートになると考えられる。

この提案手法でマージを行ったフローチャートとマージを行う前のフローチャートのノード数を比較する。またマージを行うことでフローチャートの内容に変化が起きる場合が考えられるので、フローチャートの内容に問題がないかを人手で評価する。フローチャートは、ノードはオペレータの質問、エッジはユーザの回答という形で構成されている。

第2章 関連研究

村田ら [1] は，岡崎ら [2] の研究を用い，フローチャートを効率的に作成する研究を行っていた．具体的には，情報抽出の技術を用いコールセンターでの対話データを表の形に整理し，その表からフローチャートを作成する手法を提案していた．その手法の手順を以下に示す．

2.1 表作成の手順

- (1) 対話データを質問文と回答文に分ける．
- (2) 質問文について，文中の名詞の単語ベクトルの総和を質問文ベクトルとする．
- (3) 質問文を x-means 法 [3] でクラスタリングし，表に整理する．
- (4) (3) で生成した質問文の表に対応するように，回答文の表を生成する．
- (5) (4) で生成した回答文の表について，列ごとに含まれる回答文を再度 x-means 法でクラスタリングし，対応するクラスタ番号を回答文の先頭に付与する．
- (6) 質問文の表と回答文の表の対応する列を取り出し，質問回答... の順で整理する．また，表の左右方向は列に含まれる文の出現箇所の平均が小さい順にソートする．
- (7) 5 で生成された表について，人手で修正後，回答文に付与されたクラスタ番号を元にソートをかける．

2.2 フローチャートの自動作成の手順

2.2.1 節はフローチャートを自動作成する基本の手法である。2.2.2 節から 2.2.4 節の手法は基本の手法と組み合わせて用いる補助手法である。

2.2.1 MF 基本手法

表とフローチャートが相互に変換可能である。表の階層構造を利用して、表の列がフローチャートの各段に対応する。この変換を利用して、フローチャートを自動作成する。この手法を MF 基本手法と呼ぶ。

表 2.2.1 と図 2.1 は相互に変換可能であるものの例である。表では Answer に「A1&1」といった数値が与えられる。この数値の「A 数値」の部分はその数値の番号の Question の回答であることを表し、&以降の数値は回答の種類を表す。類似する回答は&以降の数値が同じになる。

表の初めの列にある質問がフローチャートの初めの質問のノードになる。表 2.2.1 と図 2.1 では最初の質問である Question1 が 2 種類の回答で分岐している。A1&1 の回答がなされた分岐は Question2 へ続き、A1&3 の回答がなされた分岐は Question3 へと続いている。

このように表の階層とフローチャートの各段は対応しており、表とフローチャートは相互に変換が可能である。

表 2.2.1: 対話データから作成された表

	Question1	Answer1	Question2	Answer2	Question3	Answer3
対話 1	Q[134]...	A[134][A1&1]...	Q[135]...	A[135][A2&1]...		
対話 2	Q[27]...	A[27][A1&1]...	Q[28]...	A[28][A2&1]...		
対話 3	Q[59]...	A[59][A1&1]...	Q[60]...	A[60][A2&2]...		
対話 4	Q[124]...	A[124][A1&3]...			Q[125]...	A[125][A3&1]...
対話 5	Q[52]...	A[52][A1&3]...			Q[53]...	A[53][A3&2]...
対話 6	Q[40]...	A[40][A1&3]...			Q[41]...	A[41][A3&2]...

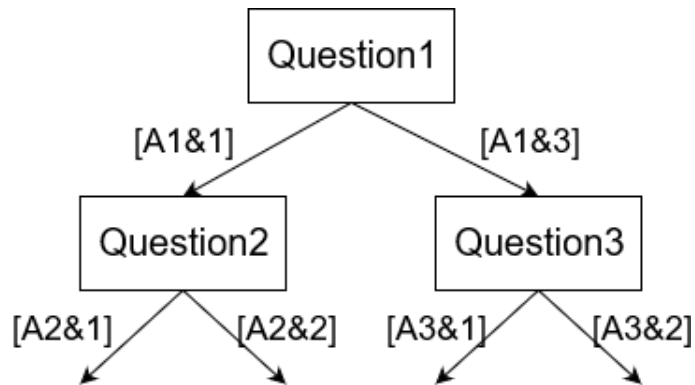


図 2.1: 表から作成されたフローチャートの例

2.2.2 MF ダミー手法

一般的に、ユーザの回答に対して、オペレータの質問は一種類に限定し、オペレータの質問に対して、ユーザは複数の回答をするので、フローチャートでは、オペレータの質問はノード、ユーザの回答はリンクで表現する。しかし、構築した表データでは、ユーザの回答に対して、オペレータが複数種類の質問をする場合があった。その場合、ダミーノードとダミーリンクを導入し、ダミーリンクから、複数の質問に対応するノードをつなげることにした。ダミーリンクとはダミーノードから出ているエッジのことである。この手法を MF ダミー手法と呼ぶ。

図 2.2 がダミーノードを含む対話の例である。ユーザの回答 1 に対してオペレータは質問 1 と質問 2 の 2 種類の質問を行うことがある。ダミーノードを追加しない場合、ユーザの回答に対して 1 つの質問しかフローチャートで表現することができない。ダミーノードを追加することによりオペレータの 2 種類の質問を含むフローチャートを表現することができる。

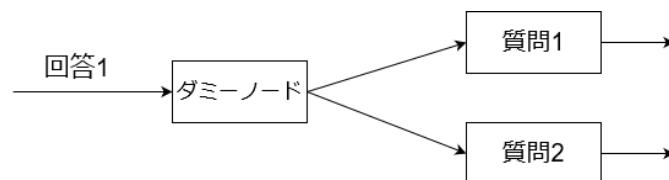


図 2.2: ダミーノードを含む対話

2.2.3 MF 単一回答削除手法

人手で修正した表であっても，表は情報が冗長な場合がある．回答の種類が一種類しかない Question と Answer の列がある場合がある．この種の質問は，オペレータは質問するが，ユーザは常に同じ回答をするもので，する必要のない質問をオペレータがしている場合が多い．例えば，映像が出ていないところでさらに音声も出ていないことを念押しで聞く場合がある．このような発話は削除しても差し支えない．そこで，回答の種類が一種類しかない Question と Answer の列を削除することにした．この手法を MF 単一回答削除手法と呼ぶ．

図 2.3 の質問 3 は回答が 2 種類あるのでこの手法では削除されない．質問 2 の回答が回答 2-1 の一つだけである場合，この手法では質問 2 と回答 2-1 は削除される．

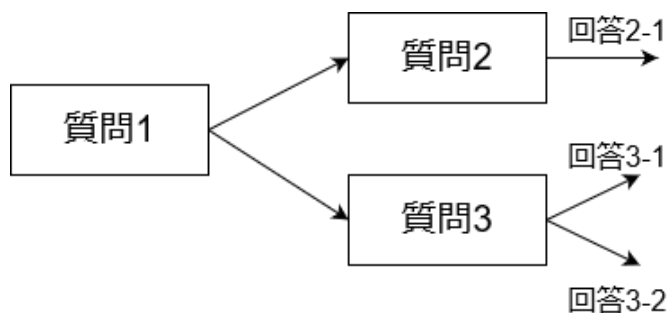


図 2.3: MF 単一回答削除手法

2.2.4 MF ダミー削除手法

ユーザの回答に対してオペレータが複数種類の発話をする場合があるが，そのうちの代表的な発話に基づくもののみでフローチャートを作成することも考えられる．複数種類の発話のうち，事例が最も多い発話のみを選択して，それ以外の発話の場合の事例を削除してフローチャートを作成する．この方法を MF ダミー削除手法と呼ぶ．フローチャートの簡単化に役立つ．

MF ダミー削除手法は，図 2.4 の質問 2 の方が出現頻度が低いため質問 2 を削除し，回答 1 から続く質問を出現頻度の高い質問 1 の一つにするという手法である．

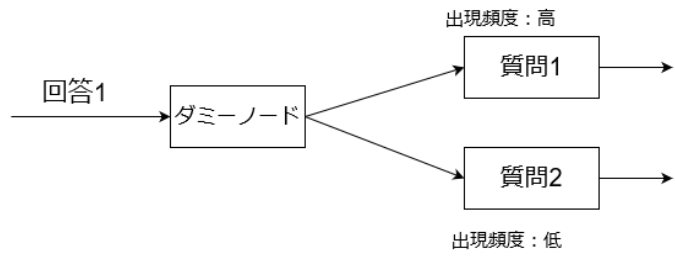


図 2.4: MF ダミー削除手法

第3章 提案手法

本研究では、対話データを自動でクラスタリングし、表の形に整理した後に、表の類似箇所のマージを行う。マージした表を用いて自動でフローチャートを作成する。

3.1 マージを行う候補

対話データから作成された表には同じ列にある質問文は内容が類似しているという特徴がある。フローチャートでの類似している質問は表で同じ列に分類されている質問のことである。この類似した質問をマージを行う際のマージ候補とする。そのマージ候補以降の対話の状況に応じてマージを行うかを判定する。

表 3.1.1 はマージを行う前の表の例である。同じ列にある質問は、その質問までになされた質問と回答の種類が異なる場合、マージを行う候補とする。また表 3.1.1 から作成したフローチャートの例が図 3.1 である。

表 3.1.1: マージ前の表

	質問 1 クラスタ 1	回答 1 クラスタ 1	質問 2 クラスタ 2	回答 2 クラスタ 2	質問 3 クラスタ 3	回答 3 クラスタ 3
対話 1	質問 1	回答 1-1	質問 2	回答 2-1		
対話 5	質問 1	回答 1-2	質問 2	回答 2-1		
対話 4	質問 1	回答 1-2	質問 2	回答 2-2		
対話 2	質問 1	回答 1-3			質問 3	回答 3-1
対話 3	質問 1	回答 1-3			質問 3	回答 3-2
対話 6	質問 1	回答 1-3			質問 3	回答 3-2

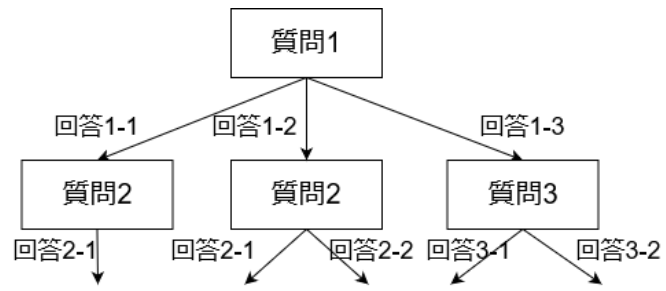


図 3.1: マージ前の表から作成されるフローチャート

3.2 マージの判定

マージ候補とした類似した質問文をそれぞれ比較することでマージを行うかを判定する。判定にはマージ候補の質問以降の対話の状態を用いる。判定方法は以下の3種類の方法である。

3.2.1 MF 同一質問ノード融合手法

MF 同一質問ノード融合手法は、フローチャートの類似する質問をマージ候補とし、マージ候補以降の対話の状況を考慮せずにマージ候補の質問は全てマージを行う手法である。類似する質問は全てマージを行うので大規模なフローチャートの規模縮小が行える。

図3.1のフローチャートでは、質問2のノードが2つ存在する。この2つを類似箇所として、MF 同一質問ノード融合手法でマージを行う。

表3.2.1は表3.1.1の類似箇所である質問2をマージした表である。対話5、対話4の質問2が対話1の質問2へとマージが行われる。MF 同一質問ノード融合手法では、マージ候補である類似した質問以降の対話の内容が異なっている場合も含めてマージを行うので、対話1と対話4は回答文が異なっているがマージを行う。●はマージ先の質問を表し、#はマージ前の質問を表している。

また、表3.2.1から作成したマージ後のフローチャートが図3.2である。

表 3.2.1: MF 同一質問ノード融合手法でのマージ例

	質問1 クラスタ1	回答1 クラスタ1	質問2 クラスタ2	回答2 クラスタ2	質問3 クラスタ3	回答3 クラスタ3
対話1	質問1	回答1-1	質問2 ●	回答2-1		
対話5	質問1	回答1-1	質問2	回答2-1		
対話4	質問1	回答1-1	質問2	回答2-2		
対話5	質問1	回答1-2	質問2 #			
対話4	質問1	回答1-2	質問2 #			
対話2	質問1	回答1-3			質問3	回答3-1
対話3	質問1	回答1-3			質問3	回答3-2
対話6	質問1	回答1-3			質問3	回答3-2

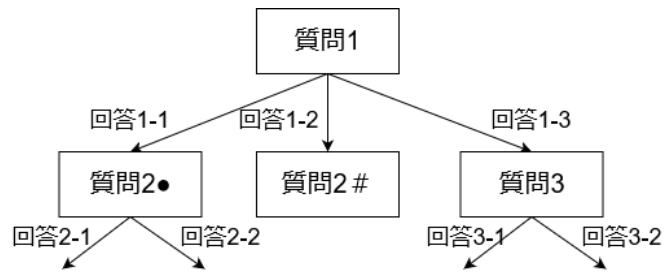


図 3.2: MF 同一質問ノード融合手法でマージしたフローチャート

3.2.2 MF 同一後続一致ノード融合手法

MF 同一後続一致ノード融合手法は、フローチャートの類似する質問をマージ候補とし、マージ候補以降の対話の内容が完全一致した場合マージを行う手法である。

表 3.2.2 と図 3.3 は完全一致している箇所を含んでいる表とフローチャートの例である。質問 1 で二つに分岐し、それぞれの分岐先で質問 3 がなされている。このとき 2 か所に出現する質問 3 の回答は、回答 3-1 と回答 3-2 の 2 つであり、2 つとも一致している。このことからこの 2 つの質問 3 は完全一致している類似箇所である。

表 3.2.2: 完全一致している箇所を含む表

	質問 1 クラスタ 1	回答 1 クラスタ 1	質問 2 クラスタ 2	回答 2 クラスタ 2	質問 3 クラスタ 3	回答 3 クラスタ 3
対話 1	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-1
対話 5	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-2
対話 4	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-2
対話 2	質問 1	回答 1-2			質問 3	回答 3-1
対話 3	質問 1	回答 1-2			質問 3	回答 3-2
対話 6	質問 1	回答 1-2			質問 3	回答 3-2

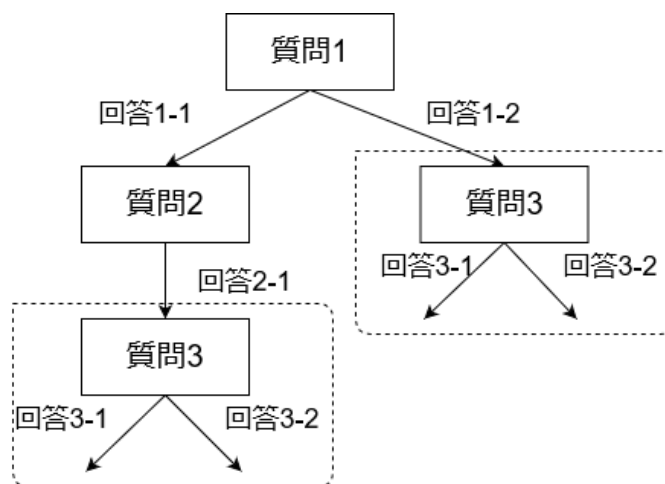


図 3.3: 完全一致している箇所を含むフローチャート

表 3.2.3 は表 3.2.2 の類似箇所が完全一致している、出現パターンの異なる質問 3 の列をマージした時の表である。図 3.3 の点線部分が類似箇所を表している。対話 2, 対話 3,

対話6で出現する質問3が、対話1, 対話5, 対話4で出現する質問3にマージされている。●はマージ先の質問を表し, #はマージ前の質問を表している。

図3.4は, 表3.2.3から作成した図3.3の完全一致している点線部分をマージしたフローチャートである。

表 3.2.3: 完全一致している箇所をマージした表

	質問1 クラスタ1	回答1 クラスタ1	質問2 クラスタ2	回答2 クラスタ2	質問3 クラスタ3	回答3 クラスタ3
対話1	質問1	回答1-1	質問2	回答2-1	質問3●	回答3-1
対話5	質問1	回答1-1	質問2	回答2-1	質問3●	回答3-2
対話4	質問1	回答1-1	質問2	回答2-1	質問3●	回答3-2
対話2	質問1	回答1-1	質問2	回答2-1	質問3	回答3-1
対話3	質問1	回答1-1	質問2	回答2-1	質問3	回答3-2
対話6	質問1	回答1-1	質問2	回答2-1	質問3	回答3-2
対話2	質問1	回答1-2			質問3#	
対話3	質問1	回答1-2			質問3#	
対話6	質問1	回答1-2			質問3#	

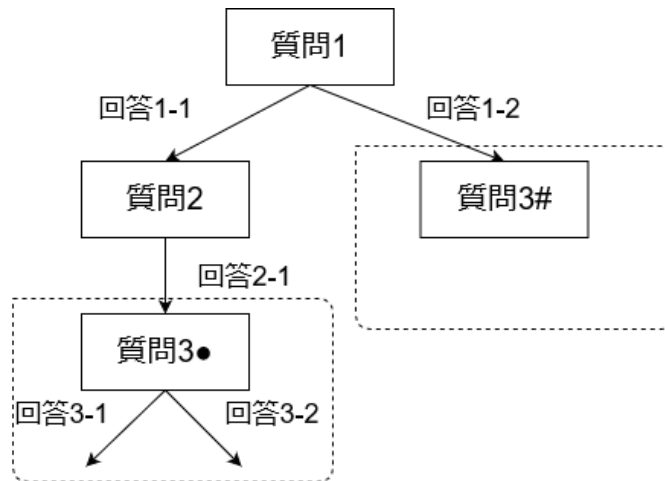


図 3.4: 完全一致している箇所をマージしたフローチャート

3.2.3 MF 同一後続類似ノード融合手法

MF 同一後続類似ノード融合手法は、フローチャートの類似する質問をマージ候補とし、マージ候補以降の対話の内容の類似度を求め、その値が0.2以上の場合マージを行う手法である。類似度の計算は、Jaccard 係数を用いて求める。Jaccard 係数 $J(A,B)$ は以下の式で定義される。

$$J(A, B) = \frac{A \cap B}{A \cup B} \quad (3.1)$$

3.1 式の集合 A, 集合 B はそれぞれマージする際に比較する、2つの類似した質問のノードから続く対話のパターン数である。

表 3.2.4 では、出現パターンの異なる質問 3 が 2 つ存在する。この 2 つの質問 3 は回答の種類が異なるので、完全一致している類似質問ではない。図 3.5 の点線部分がこのフローチャートの類似箇所である。

今回の例では、質問 2 の後に質問される質問 3 と、質問 1 の後に質問される質問 3 の 2 つである。質問 2 の後に質問される質問 3 から続く回答 3-1, 回答 3-2, 回答 3-3 の 3 種類の回答の集まりを集合 A とし、質問 1 の後に質問される質問 3 から続く回答 3-1, 回答 3-2 の 2 種類の回答の集まりを集合 B とする。このときの類似箇所の類似度は 3.1 式より、0.66 と求めることができる。このように類似度を計算し、その値が 0.2 以上の時類似箇所のマージを行う。今回の例では類似度が 0.66 であるので質問 3 をマージする。

表 3.2.4: 類似箇所の内容が異なる時の表

	質問 1 クラスタ 1	回答 1 クラスタ 1	質問 2 クラスタ 2	回答 2 クラスタ 2	質問 3 クラスタ 3	回答 3 クラスタ 3
対話 1	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-1
対話 5	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-2
対話 4	質問 1	回答 1-1	質問 2	回答 2-1	質問 3	回答 3-3
対話 2	質問 1	回答 1-2			質問 3	回答 3-1
対話 3	質問 1	回答 1-2			質問 3	回答 3-2
対話 6	質問 1	回答 1-2			質問 3	回答 3-2

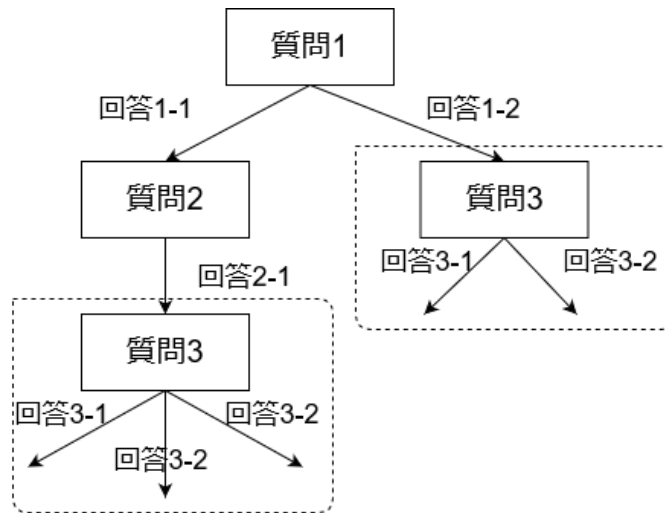


図 3.5: 類似箇所の内容が異なる時のフローチャート

表 3.2.5 は表 3.2.4 の類似箇所である，異なる出現パターンの質問 3 をマージした表である．●はマージ先の質問を表し，#はマージ前の質問を表している．

また，図 3.6 は表 3.2.5 から作成した図 3.5 の類似箇所である点線部分をマージしたフローチャートである．

表 3.2.5: MF 同一後続類似ノード融合手法でマージした表

	質問 1 クラスタ 1	回答 1 クラスタ 1	質問 2 クラスタ 2	回答 2 クラスタ 2	質問 3 クラスタ 3	回答 3 クラスタ 3
対話 1	質問 1	回答 1-1	質問 2	回答 2-1	質問 3 ●	回答 3-1
対話 5	質問 1	回答 1-1	質問 2	回答 2-1	質問 3 ●	回答 3-2
対話 4	質問 1	回答 1-1	質問 2	回答 2-1	質問 3 ●	回答 3-3
対話 2	質問 1	回答 1-2	質問 2	回答 2-1	質問 3	回答 3-1
対話 3	質問 1	回答 1-2	質問 2	回答 2-1	質問 3	回答 3-2
対話 6	質問 1	回答 1-2	質問 2	回答 2-1	質問 3	回答 3-2
対話 2	質問 1	回答 1-2			質問 3 #	
対話 3	質問 1	回答 1-2			質問 3 #	
対話 6	質問 1	回答 1-2			質問 3 #	

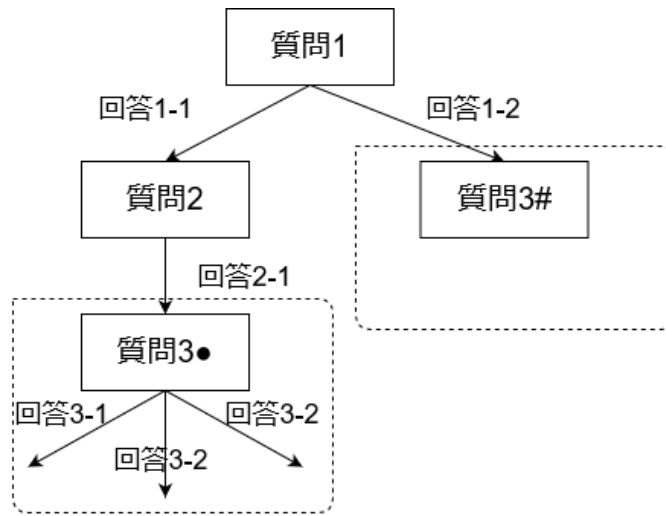


図 3.6: MF 同一後続類似ノード融合手法でマージしたフローチャート

第4章 実験

4.1 実験データ

実験には、「電源がつかない」の問診 20 件と「再生中にフリーズする」の問診 20 件を用いる。表 4.1.1 は「電源がつかない」の問診の一部である。表 4.1.1 の-はオペレータの質問，+はユーザの回答を表している。

表 4.1.1: 問診の内容 (一部)

<p>-お問い合わせありがとうございます。よろしければ問診の行わせていただきます。 〇〇の機種で電源が入らないということですが、今お近くに製品はございますか +はい、あります。</p> <p>-お手数ではありますが、確認していただきたいことが2点ございます。 レコーダ本体の電源コードはコンセントにしっかり入っていますでしょうか。 その状態でレコーダ本体側面にある電源ボタンを直接長押ししていただき 反応はありますでしょうか。 +電源コードはコンセントにしっかり入っています。 また、本体の電源ボタンを押しても反応がなく、電源ランプがつかないです。 -ご不便をかけ大変申し訳ありません。 レコーダ本体に電源ランプは点灯しておりますでしょうか。 +点灯しないです。</p> <p>-ご協力ありがとうございます。 またお手数ではありますが、お試しいただきたいことが2点ございます。 先ほど確認していただいた本体の電源ボタンを 長押ししていただき、反応はありますでしょうか。 一度電源コードをコンセントから1、2分程外していただき、 再度電源が入るか捜査をお試しいただけますでしょうか。</p>
--

4.2 評価方法

マージ前の表から作成したフローチャートとマージ後の表から作成したフローチャートのノード数を比較することで、マージによりフローチャートを簡略化できたかを評価する。このときマージした質問であることを表す記号#が質問文の文末に付与されているノードは数えないものとする。またマージによりマージ先のノードの情報量が増加し、フローチャートの内容が変化する可能性がある。フローチャートの内容の変化は、対話が良い場合と、悪くなる場合が考えられる。このことからフローチャートの内容にどのような変化が生じているかを人手で評価する。

4.3 実験結果

表 4.3.1 は手法 1, 手法 2, 手法 3 でマージを行った際のフローチャートのノード数の変化をまとめたものである。

表 4.3.1: ノード数の変化

	電源がつかない		再生中にフリーズする	
	マージ前	マージ後	マージ前	マージ後
手法 1	97	25	74	24
手法 2	97	64	74	47
手法 3	97	56	74	44

表 4.3.1 から、提案したマージ箇所の判定方法の全てで、マージを行ったことによりフローチャートのノード数が減少している。このことからフローチャートの簡略化は行えたと考えられる。次章ではマージを行うことでフローチャートの内容にどのような変化が生じたかを考察する。

第5章 考察

実験結果から、提案手法によってフローチャートを簡略化できることを確認した。

本章では、マージしたことによって起こったフロチャートの内容の変化について考察する。

5.1 マージの長所

MF 同一質問ノード融合手法でマージしたフローチャートにはマージしたことによってフローチャートの情報の増加がみられた。実際に対話に情報の増加が起こった例を図 5.1 と図 5.2 に示す。

図 5.1 の 2 種類の対話がフローチャートに出現する。この 2 種類の対話は Question6 でマージが生じ図 5.2 の対話 1 つになる。このマージが行われることによって 2 種類の対話の流れにおいて Question6 の前に Question5 が出現するようになった。またマージ先のノードである Question6 においてマージを行うことで情報が追加され分岐が増加している。

このことからマージすることによって、より詳細な対話の流れを得ることができると考えられる。また他の判定方法よりマージの回数が多いことでこのように対話内容の増加が起こる可能性が高いと考えられる。

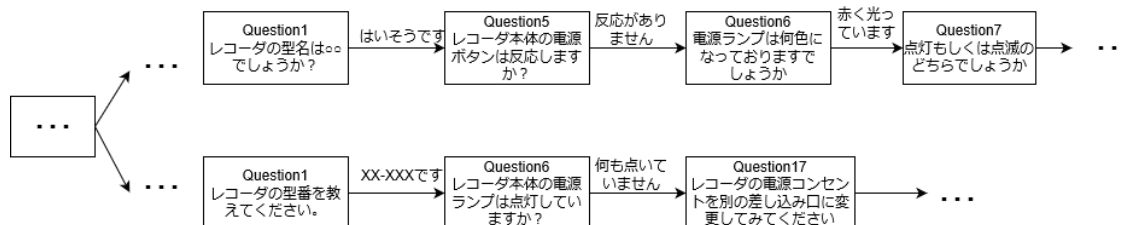


図 5.1: マージ前の対話

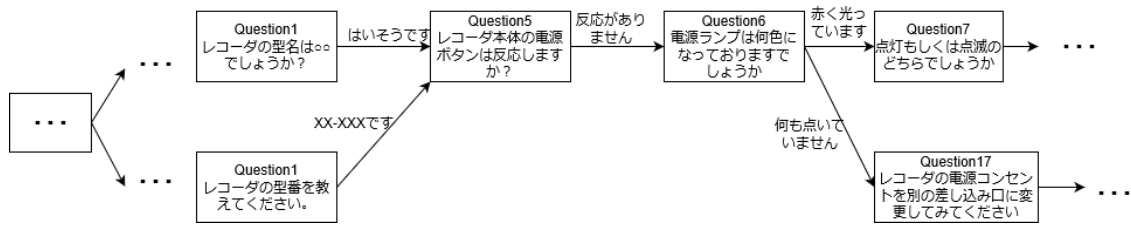


図 5.2: マージ後の情報が追加された対話

MF 同一後続一致ノード融合手法は対話の流れが完全一致した場合のみマージをするので、MF 同一質問ノード融合手法のような情報の増加は起こらなかった。しかし完全一致の場合のみマージをすることは、マージ前とマージ後でフローチャートの内容を変化させずにフローチャートの簡略化ができるという利点があると考えられる。また、対話には一度しか質問されない質問を含むユニークな対話が存在する。このようなユニークな対話がある場合必ずマージせずにフローチャートに残るという利点もあると考えられる。実際にマージが起こった対話の例を図 5.3 と図 5.4 に示す。図 5.3 は Question18 以降の対話の内容が一致している対話である。図 5.4 は Question18 でマージを行った後の対話である。

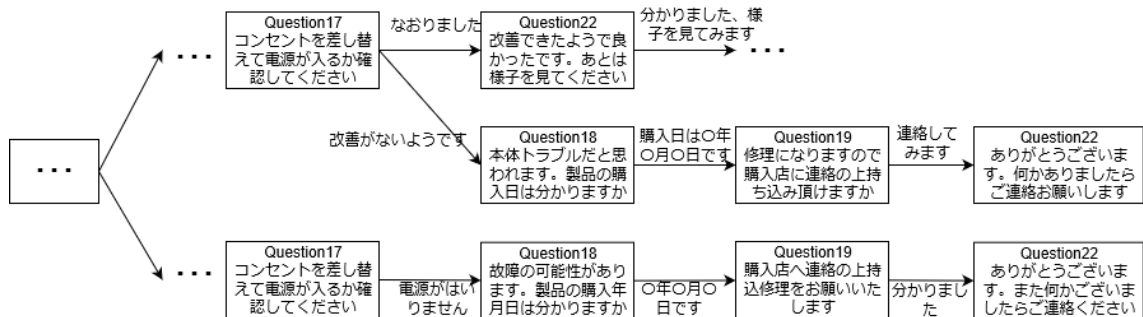


図 5.3: 完全一致している箇所を含む対話

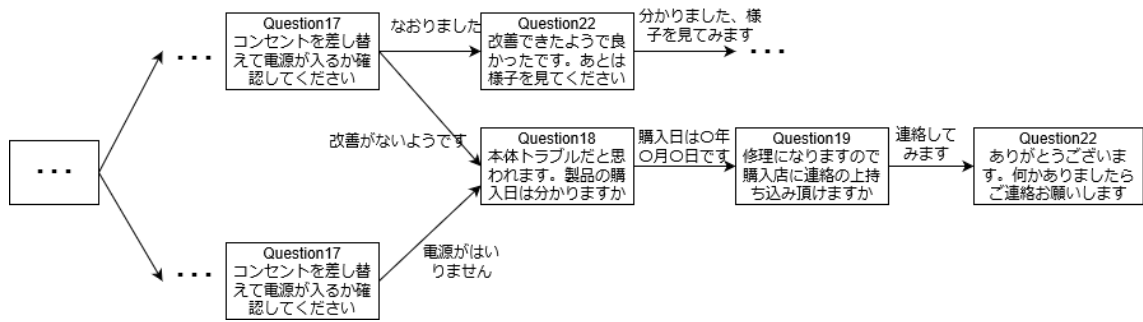


図 5.4: 完全一致している個所をマージした対話

MF 同一後続類似ノード融合手法は MF 同一後続一致ノード融合手法の制限を緩和した方法であり、対話の流れが少し異なるような場合でもマージを行うことができる。この方法は完全一致と異なり、内容に多少の変化は起こるが完全一致よりもフローチャートの規模を縮小できるという利点があった。完全一致ではマージせず、判定方法 (3) ではマージする実際の例を図 5.5 と図 5.6 に示す。図 5.5 では Question17 以降の片方の対話は内容が一致している。しかしもう片方の対話では Question19 前に Question18 が存在するかどうかの違いがある。MF 同一後続類似ノード融合手法ではこのような場合でもマージを行うことができる。図 5.6 はマージした後の対話である。

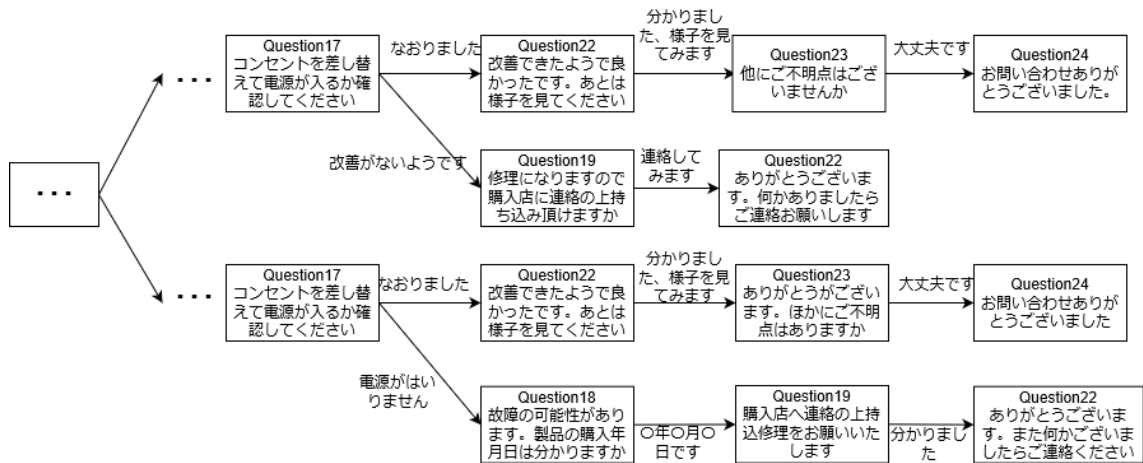


図 5.5: 完全一致ではマージしなかった対話

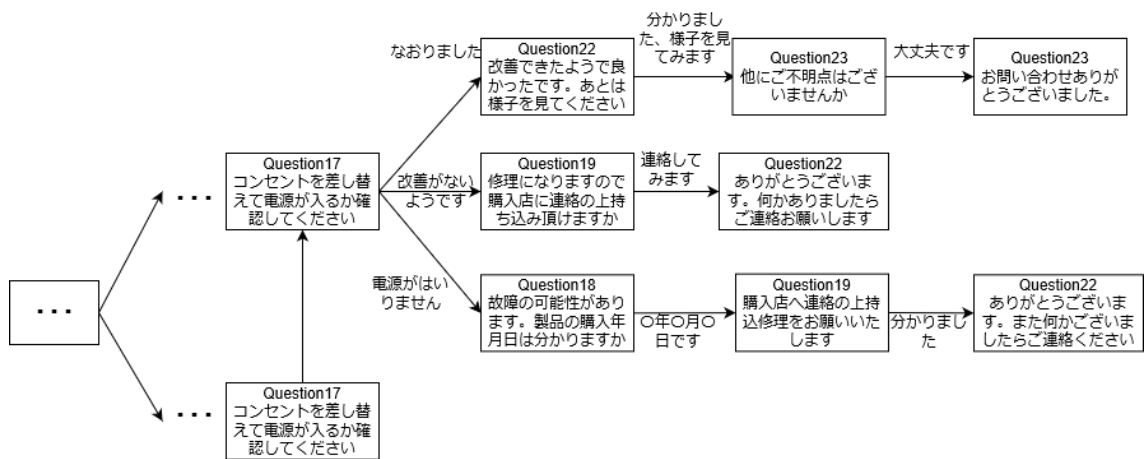


図 5.6: MF 同一後続類似ノード融合手法でマージした対話

5.2 マージの短所

MF 同一質問ノード融合手法では、マージをすることにより対話の内容が把握しづらくなる場合があった。他の対話では質問されることのないユニークな質問が含まれる対話が存在する。しかしこのユニークな質問を含む対話が、他の多くの対話に共通する質問を含んでいることで、マージが生じる。このようなユニークな質問を含む対話がマージされることで対話の流れが悪くなる場合があった。図 5.8 の Question19 から Question20 への流れが実際に対話の流れが悪くなった実例である。

マージ前は Question19 で修理申し込みをし、Question20 でお客様情報の確認を取り、Question21 で日程調整を行う流れであった。しかしマージをしたことによって Question19 で購入店に連絡を提案し、Question20 でお客様情報の確認を取り、Question21 で日程調整を行う流れになり、対話が不自然になった。

対話の流れの大部分がユニークなものはマージせず、一本の状態でもフローチャートに存在するとより素早く内容を把握できると考えられる。またこのような対話はマージすることで対話が不自然になるので、マージをするべきではない対話であると考えられる。このことから MF 同一質問ノード融合手法の短所としてユニークな対話をマージしてしまう可能性があるという問題点が考えられる。実際にユニークな対話をマージした例を図 5.7 と図 5.8 に示す。図 5.7 の Question20 と Question21 はこの対話でしか質問されないユニークな質問である。図 5.8 はユニークな質問を含む対話をマージしてしまった時の対話である。

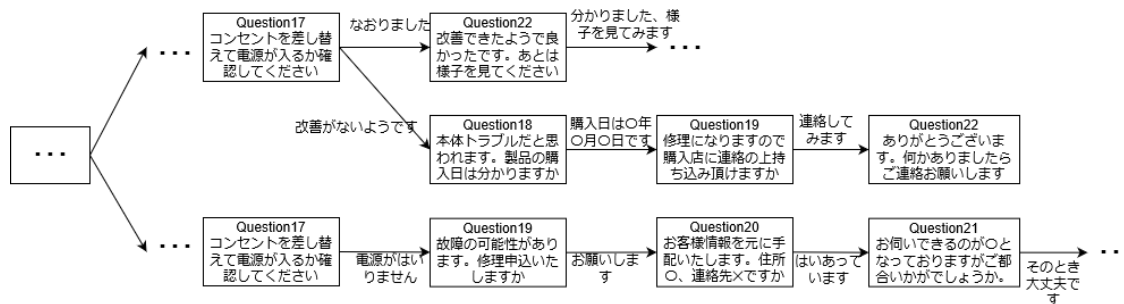


図 5.7: マージ前のユニークな対話

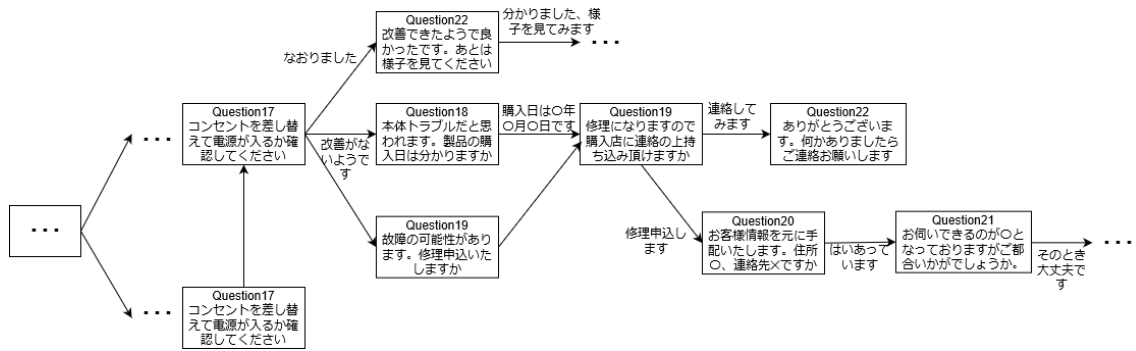


図 5.8: ユニークな対話をマージした対話

MF 同一後続一致ノード融合手法では完全一致のみマージをするので、対話の流れの大部分が一致していても対話の一つ質問が多い場合など、対話の内容が一部でも異なるだけでマージが行われずに類似した内容がフローチャートに残る場合があった。このようなものがマージされずに残ってしまうことが MF 同一後続一致ノード融合手法の短所であると考えられる。実際に残ってしまう対話の例を図 5.9 に示す。図 5.9 には 2 つ Question17 出現するが、片方の Question17 は Question22 と Question19 へと続き、もう片方の Question17 は Question22 と Question18 へと続いており、Question17 以降の対話が一致していない。このことから 2 つの Question17 はマージされずにフローチャートに残ってしまう。この 2 つの対話の違いは片方は Question17, Question18, Question19 という順番で質問されるか、Question17, Question19 という順番で質問されるかの差である。Question18 は購入日を聞く質問であるが今回の対話の流れでは、質問がない場合でも対話内容にあまり変化はないと考えられるので、この 2 つの対話はマージしても良いと考えられる。

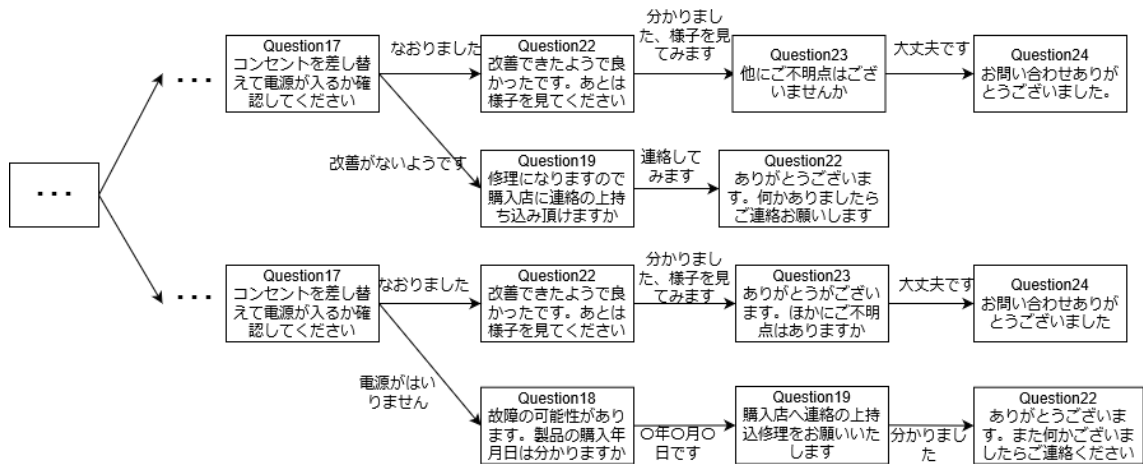


図 5.9: MF 同一後続一致ノード融合手法ではマージできない対話

MF 同一後続類似ノード融合手法では長所で上げたような対話の流れが少し異なるような場合でもマージを行う。しかしこの方法の短所として、異なっている個所が他の対話にはないユニークな質問や回答であった場合、マージをすることで不自然な対話になる可能性があるということが考えられる。

また全ての方法において、表で質問の表現は似ているが内容が異なる質問が同じ質問として扱われることにより、マージが行われ対話の内容が悪くなる場合があった。これは最終的な対応などが異なる場合があるのでマージすべきではないものがマージされ

てしまった事例である。実際の例を図 5.10 と図 5.11 に示す。

図 5.10 では、Question22 が 2 つ出現するが質問の内容がそれぞれ「様子を見てください」と「修理受付が完了した」であり内容が異なっているが同じ Question22 に分類されていることでマージされてしまう。このような箇所をマージしてしまうことで対話の内容が悪くなることがあった。

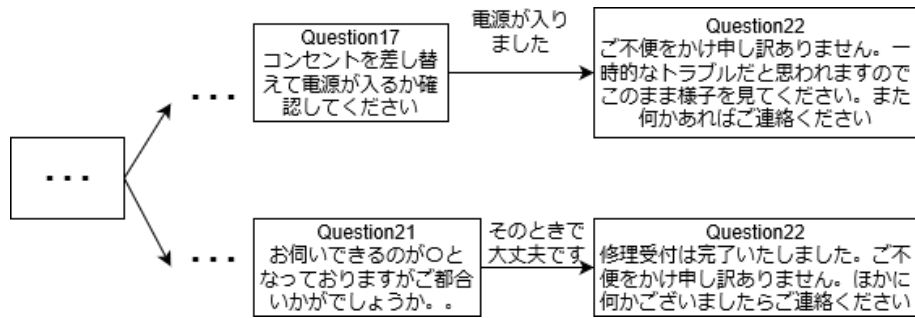


図 5.10: 異なる内容の質問が類似質問となされている対話

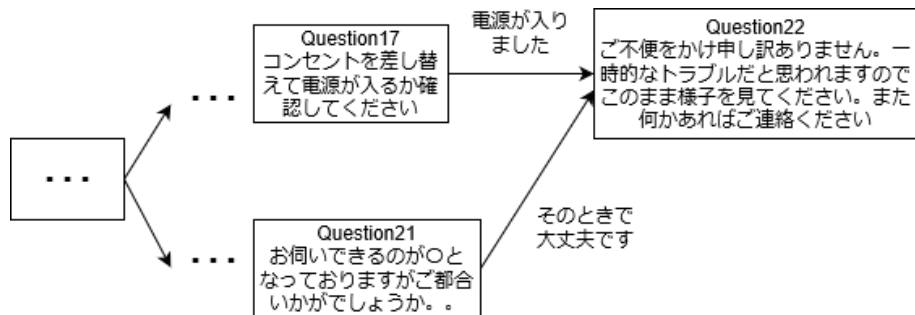


図 5.11: 異なる内容である質問がマージされた対話

第6章 今後の課題

5章の考察からフローチャートはマージを行うことで内容が変化することがあるということがわかった。しかし内容が変化し対話が不自然になってしまう場合があった。MF同一後続一致ノード融合手法は内容が変化することは考えられないが、他の判定方法と比較して、フローチャートの規模の縮小の幅が小さい。

そこで今後の課題として、フローチャートの対話が不自然なものにならずに、フローチャートの簡略化が行えるようなマージの判定基準を見つける必要がある。

表では同じ列には内容が類似した質問が分類分けされているが、実際の表では表現は似ているが内容が異なる質問が同じ質問として分類分けされることがあった。このような分類がなされた質問がマージされ対話の流れが悪くなることがあった。このときの例として図5.10と図5.11がある。このことから表の整理を行う際、どの程度の質問を類似した質問とするとよりよいマージが行えるのかも検討していきたい。

第7章 おわりに

本研究では、類似箇所が出現するフローチャートを簡略化する手法を提案した。

提案手法は、類似箇所をマージすることによって、フローチャートの規模を縮小し、フローチャートの簡略化を行った。マージの判定は類似箇所の対話の内容を基準に3種類の判定方法で行った。

「電源が見つからない」の問診20件と「再生中にフリーズする」の問診20件で提案手法の3種類の判定方法で実験を行った結果、「電源が見つからない」のフローチャートのノード数が97個からそれぞれ25個、64個、56個に減少し、「再生中にフリーズする」のフローチャートのノード数が74個からそれぞれ24個、47個、44個に減少した。3種類のマージの判定方法全てでフローチャートのノード数が減少し、フローチャートの規模が縮小した。

フローチャートの簡略化は行えたが、マージをする際にフローチャートの対話内容に変化が起こることがあった。情報が増加しより良い内容になることもあったが、内容が変化し対話の流れが不自然になる場合もあった。類似箇所をマージしているためこのようなことが起きる。

今後の課題として、類似箇所でもマージを行っても対話が不自然にならないマージ箇所の判定方法を見つけることが課題である。

謝辞

最後に，1年間に渡りご指導いただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村田真樹教授，村上仁一准教授に心から御礼申し上げます．そのほか様々な場面でご助言していただいた自然言語処理研究室の皆様へ感謝の意を表明します．

参考文献

- [1] 村田真樹, 守優太郎, 小山一人, “表整理技術を用いたコールセンター疑似対話データからのフローチャート作成” 電子情報通信学会信学技報, NLC2019-1, pp.1-6, 2019
- [2] 岡崎健介, 村田真樹, 馬青, “複数文書からの重要情報の抽出と表の作成” 言語処理学会第 24 回年次大会, pp.240-243, 2018
- [3] 石垣恒憲, “x-means 法改良の一提案: k-means 法の逐次繰り返しとクラスターの再併合” 計算機統計学, 第 18 巻, pp.3-13, 2006