

ニュース関連ツイートの感情極性に基づく 共感対話システムの提案と評価

田原 俊一[†] 池田 和史[†] 松本 一則[†] 帆足啓一郎[†]

[†] KDDI 総合研究所 356-8502 埼玉県ふじみ野市大原 2 丁目 1 番 15 号

E-mail: †{sh-tahara,kz-ikededa,matsu,hoashi}@kddi-research.jp

あらまし 対話システムにおいて、ユーザの発話内容に対しシステムが共感を示す応答を行うことは、対話を活性化させユーザの満足を向上する。ユーザに共感を与えるシステムの応答として相槌等があるが、応答の種類数が乏しく対話が終了しがちになるため、バリエーション豊かでユーザの発話内容の感情を考慮した応答が可能な手法が必要となる。本稿では、ユーザの発話の極性に応じて、ニュースに関連するツイートを選択して応答する対話システムを提案する。予備実験を通して、ユーザの発話と同じ極性 (Positive, Negative) のツイートをシステムの応答とすることで、ユーザの共感が高まることが示唆された。極性を考慮した提案対話手法と既存の対話手法の比較評価実験を行い、提案手法は既存手法と比較し有効性を確認した。

キーワード 対話システム, HCI, 共感, SNS, 感情分析

1. ま え が き

近年、日本マイクロソフト社が開発したりんな [1] といったユーザと雑談を行う対話システムが広く利用されている。ユーザの発話 (以下、ユーザ発話と定義) に対しシステムが共感を示す応答をすることは、ユーザの満足を向上する可能性がある。例えば、Dodge [2] らは映画に関する対話実験を実施しており、映画の面白さに言及するユーザ発話に対し、システムは共感を示す応答をしている。

山口ら [3] は、学生とカウンセラーとの対話において、学生の発話とカウンセラーの相槌を学習させ、発話を入力した際に適切な相槌を予測する手法を提案している。石田ら [4] は、ユーザ発話に対して同意や繰り返し応答等を行う手法を提案している。これらの手法では、システムの応答 (以下、システム応答と定義) が一定数であるため応答の種類数が乏しいがユーザの共感度を向上する提案を行っている。Jiwei ら [5] によれば、過度に学習した Deep Learning を用いた対話システムでは、「分からない」といった種類数が乏しい応答をしがちになり、その結果対話が終了してしまうことが報告されている。より共感度の高い対話システムの実現のために、バリエーション豊かでユーザの感情を考慮した応答が可能な手法が必要となる。

本稿では、ユーザ発話の極性に応じて、ニュース等のコンテンツに関連するツイートを選択して応答する対話システムを提案する。ここで、ユーザ発話の極性に応じてツイートを選択することが最も共感を生む選択であるかは自明ではない。そのため、ユーザ発話の極性に対して同じ極性のツイートを応答すればユーザの共感度は高くなるという仮説を立て検証を行う。

提案手法の有効性を評価するため、ユーザ発話に対し相槌を応答する手法、ニュース記事に関連するツイートからランダムにツイートを選択し応答する手法と比較し、評価実験を実施した。具体的には、被験者とシステム間で各手法に沿って対話を

行い、各対話に対する被験者の共感度をアンケートを用いて評価する。評価の結果、提案手法の平均共感度が有意に高いことが確認された。

本稿の構成は以下の通りである。2 章で関連研究について述べる。3 章では提案手法を述べ、4 章では予備実験について述べる。5 章では提案手法の評価について述べる。最後に 6 章で結論と今後の課題について述べる。

2. 関連研究

ユーザ発話に対し、感情を含む応答を行う対話システムの研究は様々行われている。Zhou ら [6] は、ユーザの投稿と、6 つの感情 (好き, 幸福, 悲しい, 嫌悪, 怒り, その他の感情) の内 1 感情をアノテーションした応答文を学習し、分類器にユーザの投稿を入力すると、6 つのそれぞれの感情が含まれている 6 つの応答を出力するシステムを提案している。このシステムでは中国語の会話データを学習し、応答を生成している。

これまで、ユーザの発話に対して共感を示す応答を行う対話システムに関する様々な研究が行われてきた。山口ら [3] は、学生とカウンセラーとの対話において、学生の発話とカウンセラーによる相槌を学習させ、発話が入力された際、適切な相槌を予測する手法を提案している。音声を用いた評価実験では、適切な相槌を予測することで被験者の共感度が上がっている。石田ら [4] は、ユーザ発話に対して「そうですか」といった同意を示す応答や、ユーザ発話の一部分を繰り返して応答を行う手法を提案している。これらの手法では、応答は一定数であるため応答の種類数が少ないが、ユーザの共感度を向上する提案を行っている。一方で、Jiwei ら [5] によれば、過度に学習した Deep Learning を用いた対話システムでは、「分からない」等の種類数が乏しい一般的な応答をしがちになってしまい、その結果対話が終了してしまうことが報告されており、良い応答ではないと主張している。従って、より共感度の高い対話システ

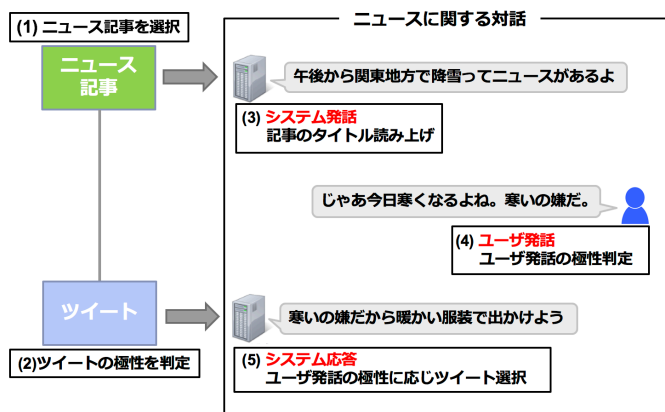


図1 提案手法における対話システムの構成

ムを実現するためには、バリエーションが豊かでユーザの感情を考慮した応答が可能な手法が必要となる。

3. 提案手法

2章で述べた課題に対して、本稿ではツイートボタンが設置されているウェブニュースの記事とボタンを通して生成されたツイートを取得し、ユーザ発話の極性に応じてツイートを選択して応答する対話システムを提案する。ウェブニュースの記事といったコンテンツの話題に限定したツイートを応答することで、コンテンツに基づく内容の濃い対話が可能になり、ユーザ発話の極性と同じツイートを応答することでユーザの共感性が高くなると思われる。

提案手法における対話システムの構成を図1に示す。システムは次の手順でユーザとの対話を行う。(1) コンテンツ情報をウェブニュース記事とする。ユーザが登録しているTwitterアカウントのプロフィール情報や、ユーザが投稿したツイートから趣味嗜好に関する情報を取得し、その情報に関連するニュース記事をニュース配信サイトから1件選択する。(2) 選択したニュース記事に関連するツイート群を取得し、Google Cloud Platform [7] が提供する Natural Language API の感情分析 (Google API) を用いて各ツイートを Positive・Negative・Neutral の3極性に分類する。(3) その後、ニュースに関連する対話を開始し、記事のタイトルをシステムが読み上げる。本稿では、これをシステム発話と定義する。(4) システム発話に対してユーザが任意の発話を行い、Google API を用いて発話内容の極性の判定を行う。(5) ユーザ発話の極性に応じて、(2) で3極性に分類したツイート群の中から1つを選択し、その群の中で特に主観表現を豊富に含むツイート1件を文献[8]の手法を用いて選択する。その後、また(4)に戻りユーザが発話を行う。

4. 予備実験

3章で述べた提案手法において、ユーザ発話の極性に応じてツイートを選択することが最も共感を生む選択であるかは自明ではないことが問題として挙げられる。問題を解決するため、

ユーザ発話の極性を考慮した応答をどのように選ぶべきか明らかにする必要がある。文献[9]によると、共感とは相手が抱いている感情と同様の感情を自分自身も抱いていることを示すことであり、相手と同じ感情を示すことは共感を生み出す方法として重要であると述べている。そのため、ユーザ発話の極性と同じツイートをを選択すると、ユーザの共感性が高くなると考えられる。ユーザ発話の極性に対してシステムが同じ極性のツイートを応答すれば、ユーザの共感性は高くなるという仮説を立て検証を行う。例えば、ユーザの Negative な発話に対し、システムは Negative に応答する場合と Positive に応答する場合で、どちらがユーザに共感を与えるかを調べる必要がある。図1の(2)と(5)において、ツイートを手動で選択して複数の発話からなる対話データ(以下、実験データと定義)を作成し、各対話に対する被験者の共感性をアンケートを用いて評価する。本章では、実験データの収集、評価方法、評価結果に関して述べる。

4.1 実験データの収集

2017/7/1~11/10に配信されたYahoo!ニュースの記事の中から、被験者がニュースタイトルを見ただけで何を意図しているニュースなのか把握することが可能なニュース記事3件を評価に利用した。

4.1.1 ツイートの取得

一般的に、ニュース記事に関連するツイートの中には、(1)URLを含むツイートや、(2)本文の文字数が多いツイート、(3)「!!!!」のように連続した文字を含むツイートが散見される。(1)~(3)のツイートをシステムの応答とすると、ユーザが長い文章を読むため時間が掛かる。その結果、ユーザが不快に感じてしまうため、(1)~(3)のツイートは対話の内容として適さない。従って、収集したツイートの内、(1)URLを含むツイート、(2)本文が61文字以上であるツイートを除外[10]する。更に、(3)ツイート中に含まれる4文字以上の連続した文字列を3文字まで削る。

4.1.2 ツイートの極性決定

4名の作業員でツイートに極性を付与し、Positive, Negative, Neutral, 感情なしの4値でアノテーションを行う。ここで、Neutralは「有給休暇を紙で申請するのが気になって仕方ない」といったPositiveでもNegativeでもない感情を示しており、感情なしは「有給休暇はいかなる理由でも取得可能だよ」といったように何の感情も含まないことを示している。本実験では仮説の検証を行っており、ユーザ発話とツイートの感情極性に着目しているため、感情を含まないツイートは対象外としている。4名中3名以上が1ツイートに対して同じ極性にアノテーションした場合、そのツイートの極性を3名以上がアノテーションした極性に決定する。例えば、あるツイートに関して3名がPositive、1名がNeutralにアノテーションした場合、そのツイートの極性をPositiveに決定する。4名中3名以上が1ツイートに対して感情なしにアノテーションした場合、そのツイートを除外する。

4.1.3 対話の生成

システム発話として記事のタイトルを提示し、ユーザ発話及びシステム応答としてツイートをそれぞれアノテーションを行っ

表 1 実験データ

記事の要約	システム発話	ユーザ発話	システム応答	対話数
<要約文>	君が 興味ありそうな ニュースだよ。 Yahoo!ニュース によると <タイトル> らしいよ。	Positive	Positive	5
		Positive	Negative	5
		Positive	Neutral	5
		Negative	Positive	5
		Negative	Negative	5
		Negative	Neutral	5
		Neutral	Positive	5
		Neutral	Negative	5
	Neutral	Neutral	5	

表 2 共感度と評価項目

共感度	評価項目
-3	全く感じさせない
-2	ほとんど感じさせない
-1	どちらかというと感じさせない
0	どちらともいえない
1	どちらかというと感じさせる
2	ほとんど感じさせる
3	非常に感じさせる

た作業者とは別の被験者に提示する。表 1 のように、ユーザ発話及びシステム応答それぞれに 3 つの極性を割り当て、計 9 条件を設定する。各条件ごとに 5 対話を用意し、計 45 対話分の対話データを用意する。尚、表 1 の Positive は極性が Positive に決定したツイート、Negative は極性が Negative に決定したツイート、Neutral は極性が Neutral に決定したツイートを示す。被験者に実験データを提示する際に、表 1 の 45 対話分をランダムに並びかえる。被験者にニュースの内容を正しく理解してもらうため、実験データの中に記事の要約も記す。

4.2 評価方法

4.1 節で生成した実験データの各対話に対して、「ユーザ発話に対するシステム応答がユーザに共感を感じさせたと思うか」という問い方で、被験者 8 名に対し -3 から 3 までの共感度のアンケートに回答させる。共感度と評価項目は表 2 に示す。

4.3 評価結果

ニュース記事 3 記事分における、被験者 8 名の全対話の平均共感度が表 3 である。ここでは、平均共感度が 0 よりも大きい場合、共感が得られたとみなすものとする。ユーザ発話が Positive の場合、平均共感度が最も高いシステム応答の極性は Positive であり、応答として Positive を示すと共感が得られることが分かった。また、ユーザ発話が Negative の場合、平均共感度が最も高いシステム応答の極性は Negative であり、応答として Negative を示すと共感が得られることが分かった。ユーザ発話が Neutral の場合、平均共感度が最も高いシステム応答の極性は Neutral であったが、どの極性を応答してもユーザの共感は得られないことが分かった。

また、ユーザ発話の 3 極性それぞれにおける、システム応答の 3 極性間の有意差を Kruskal-Wallis 検定により分析した。ユーザ発話の極性が Positive の場合、Kruskal-Wallis 検定によって

表 3 予備実験における平均共感度

		システム応答		
		Positive	Negative	Neutral
ユーザ発話	Positive	0.983	-1.217	-0.417
	Negative	-0.908	0.058	-0.583
	Neutral	-0.417	-0.583	-0.158

有意水準 1% でシステム応答の極性が Positive と Negative 間、Positive と Neutral 間に有意差が認められた。また、同検定によって有意水準 5% でシステム応答の極性が Negative と Neutral 間に有意差が認められた。ユーザ発話の極性が Negative の場合、同検定によって有意水準 1% でシステム応答の極性が Positive と Negative 間に有意差が認められた。ユーザ発話の極性が Neutral の場合は有意差が認められなかった。

5. 提案手法の評価

4 章で得られた結果から、ユーザ発話が Positive または Negative であれば同じ極性のツイートを応答に利用し、ユーザ発話が Neutral であれば、文献 [3] を参考に相槌を応答とする手法を 3 章で述べた提案手法に導入する。

図 1 における (1)~(3) の手順を踏んだ後、(4) でユーザ発話の極性が Positive または Negative と判定されたら、(5) ではユーザ発話と同じ極性のツイート群から文献 [8] の手法を用いて 1 件のツイートを選択する。(4) でユーザ発話の極性が Neutral と判定されたら、(5) では「うん」、「うんうん」、「うんうんうん」、「そうだね」、「そうね」、「確かに」という 6 種類の相槌からランダムで 1 つ応答を行う。

提案手法の有効性を評価するため、ユーザ発話に対しシステムが 6 種類の相槌からランダムで 1 つ応答を行う相槌手法、システムがニュース記事に関連するツイートをランダムに 1 件選択し応答とするランダム手法と比較し、評価実験を実施した。3 手法に沿って対話を実施し、システムの応答に対して被験者は共感を得られたかどうかアンケートを用いて提案手法の評価を行う。本章では、評価方法及び評価結果に関して述べる。

5.1 評価方法

5.1.1 実験環境

システム発話→ユーザ発話→システム応答→ユーザ発話→システム応答の計 5 発話を 1 対話とする。複数のニュース記事を用意し、ニュース内容に関する対話を被験者とシステム間で行う。実験は、対話システムのアプリケーションを搭載したスマートフォンを用いて行う。ユーザがシステムと対話するための手順を説明する。図 2 はシステムの画面を示している。(1) はシステム発話であり、ニュース記事のタイトルをシステムが発話する。(2) はシステム発話に対するユーザ発話であり、図 2 の下部にある入力フォームを用いて被験者がテキストで入力を行う。(3) はユーザ発話に対するシステム応答であり、各手法に沿って応答を行う。(4) はシステム応答に対するユーザ発話であり、(5) はそのユーザ発話に対するシステム応答である。

5.1.2 ニュース記事及びツイートの準備

ランダム手法、提案手法ではニュース記事に関連するツイ

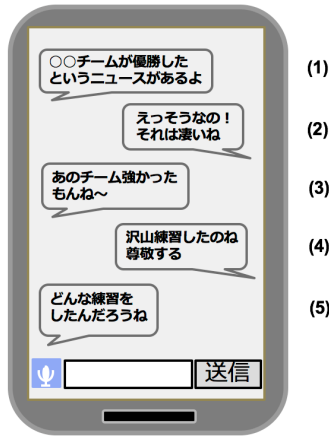


図 2 対話システムのアプリケーション

トをシステムの応答としている。2017/11/1 ~ 11/28 に配信された Yahoo!ニュースの記事を取得し、これらのニュース記事に関連する引用ツイート数の多い順に記事を並び替え、上位 30 記事の中から 3 記事を取得し実験に使用する。その際、被験者がニュースタイトルをみただけで何を意図しているニュースなのか把握することが可能なニュース記事を取得する。ツイートに関しては、ニュース記事の内容に関連性が高いツイートを取得するため、3 記事に関連している引用ツイート計 1033 件を取得した。対話に適した文を抽出するため、その中から Twitter ユーザが書き込みを行なったコメント以外（例えば、URL や記事のタイトル名）を除去する。

5.1.3 提案手法における極性判定

提案手法では、ユーザ発話と、5.1.2 項で集めたツイートの極性判定を Google API を用いて実施する。実験開始前にツイートを Google API に入力し、極性の判定結果に基づき 1 記事毎に、Positive, Negative, Neutral の 3 極性にツイートを分類する。実験開始後、入力されたユーザ発話を Google API に送信し、3 極性の判定を行う。ユーザ発話が Positive 若しくは Negative と判定された場合、それぞれ Positive, Negative であるツイート群を選択し、そのツイート群からユーザ発話に対する応答として相応しいツイートを 1 件選択し、システム応答とする。ユーザ発話が Neutral と判定された場合、6 種類の相槌から 1 つ応答する。

5.1.4 評価項目

8 名の被験者が実験に参加し、各被験者は相槌手法・ランダム手法・提案手法の順にそれぞれニュース記事 1~3 に関する対話を行い、評価を行う。各対話を実施後、「あなたの発話に対するシステムの応答に対して、共感を感じましたか?」という問い方で、4.2 節の表 2 と同じ共感度及び評価項目で、アンケートに回答させる。アンケートでは「相槌手法」「ランダム手法」「提案手法」をそれぞれ「A 方式」「B 方式」「C 方式」と匿名化した。全対話が終了した後に、「A~C の方式の内、どの方式が良かったと思えましたか?」、「A~C の方式の内、どの方式が悪かったと思えましたか?」と問い、自由記述による主観評価を求める。

表 4 各手法における平均共感度

	相槌手法	ランダム手法	提案手法
平均共感度	-0.125	0.000	1.000
分散	2.943	2.417	1.667
中央値	0	0.5	1

表 5 手法の良し悪しに関連する被験者の人数分布

	相槌手法	ランダム手法	提案手法
良かった手法	2(名)	1(名)	4(名)
悪かった手法	4(名)	3(名)	1(名)

5.2 評価結果

被験者 8 名の、各手法毎のニュース記事 1~3 における全対話の平均共感度が表 4 である。ここでも予備実験と同様に、平均共感度が 0 よりも大きい場合、共感が得られたとみなすものとする。実験結果より、3 手法の中で平均共感度が最も低い手法が相槌手法で、最も高い手法が提案手法であることが分かった。また、平均共感度が 0（どちらともいえない）を上回った手法は提案手法のみである。

共感度に関する有意差を Friedman 検定により分析した。Friedman 検定によって有意水準 5% で、(1) 相槌手法と提案手法の間、(2) ランダム手法と提案手法の間で有意差が認められた。具体的には、(1) では自由度 2 のカイ二乗分布で統計量が 9.000 であるため有意水準 5% で有意差が認められ、(2) では自由度 2 のカイ二乗分布で統計量が 7.563 であるため有意水準 5% で有意差が認められた。

表 5 は、3 手法の中でどの手法が最も良かったか、又は悪かったかという質問項目に関する被験者の人数分布を示している。良かった手法の中で最も人数が多かった手法は提案手法で、悪かった手法の中で最も人数が多かった手法は相槌手法であることが分かった。

実験の結果、提案手法は表 4 で示したように 3 手法の中で平均共感度が最も高く、表 5 で示したように最も良かった手法と回答した被験者が多かった。

5.3 考察

表 6 は各手法毎の共感度の高かった対話例と低かった対話例を示している。表 7 は、3 手法の内どの手法が最も良かったか、又は最も悪かったかという質問項目に関する被験者のコメントを示している。

相槌手法に関しては、「当たり障りのない返答だが会話は成立している」「相槌だけではあるが、比較的自然的な対話」というポジティブな意見がある一方で、「回答が面白くない」、「共感できない」、「意味を感じられない」、「相槌しかない」というネガティブな意見が多かった。例えば、表 6 のニュース A の相槌手法における共感度の低かった対話例のように、「他のスポンサーは?」というユーザの質問に対し、システムは「確かに」と応答しており、質問に対して答えられていない。

ランダム手法に関しては、「応答が一方向的に感じた」、「3 方式の中で最も共感を得られない」というネガティブな意見が見られ、対話のログを分析した所、ユーザ発話に対して文脈に合わ

表 6 各手法毎の共感度の高かった対話例と低かった対話例

ニュース記事	手法	共感度の高低	ユーザ発話	システム応答
ニュース A 「東芝がサザエさん 降板へ 放送開始から 48 年間スポンサー」	相槌	共感度が高い	そんなに長く続いてたんだね	うんうん
		共感度が低い	他のスポンサーは？	確かに
	ランダム	共感度が高い	悲しいね	サザエさんごと終わってしまうのかのような寂しさ
		共感度が低い	次のスポンサーは誰になるのでしょうか	じゃあ、来年の春からは加藤さんの OP スポンサーコールが無くなるのか（泣）
	提案	共感度が高い	サザエさん、小さいときから見てたなー	時代の流れを感じる記事だ。
		共感度が低い	-	-
ニュース B 「東京ディズニー ランドで労災認定 着ぐるみで腕に激痛」	相槌	共感度が高い	着ぐるみの人も大変そうだね	うんうん
		共感度が低い	労災なんだ	そうだね
	ランダム	共感度が高い	きぐるみ大変そう	着ぐるみ言うなー（笑）
		共感度が低い	労災なんてどこでもあるよね	えっ！ミッキーとかミニーって着ぐるみだったの!?
	提案	共感度が高い	もっと人増やしてむりのないように していかないとね	中の人などいない
		共感度が低い	労災なんてどこでもあるよね 着ぐるみ大変そう	「着ぐるみ」!?! そんなわけないよ！ そうだね
ニュース C 「休日の外出離れ 顕著に -20 代男性、 ネット普及で- 国交省調査」	相槌	共感度が高い	やっぱり自宅が一番だよ	うんうんうん
		共感度が低い	寝ていたい	うんうんうん
	ランダム	共感度が高い	わかる	買い物も遊ぶのもネット・通話で済むし、 外に出ないわなあ
		共感度が低い	お金がないもん	まあこれはネットは無関係ちゃうやろなあ。 それとて金が無いということと無関係じゃないが。
	提案	共感度が高い	確かに休みの日は家で ゴロゴロしていたいよな	若者のなんとか離れシリーズも つい到这里までいってしまったのか…
		共感度が低い	お金がないからね	わかるわー。外出面倒だもんね

表 7 手法の良し悪しに関連する被験者のコメント

	相槌手法	ランダム手法	提案手法
良かった手法	<ul style="list-style-type: none"> ・ 当たり障りのない返答だが成立している ・ 相槌だけではあるが、比較的自然的対話 	<ul style="list-style-type: none"> ・ 自分の発話に対し反応がある 	<ul style="list-style-type: none"> ・ 比較的意味ある回答が多い ・ 会話として成立している ・ あまり違和感がない ・ 基本的に相槌を打ち共感を感じた
悪かった手法	<ul style="list-style-type: none"> ・ 回答が面白くない ・ 共感できない ・ 意味を感じられない ・ 相槌しかない 	<ul style="list-style-type: none"> ・ 応答が一方向的に感じた ・ 3 つの中で一番共感を感じない 	<ul style="list-style-type: none"> ・ 会話に関係ない返事が多い

ない応答をしている対話が多く見られた。表 6 の各ニュースのランダム手法における共感度の低かった対話例にもあるように、ユーザの発話に対し、全く関係のない応答をしている。一方で、「自分の発話に対して反応がある」という意見もあり、ログを分析した所、ユーザ発話に対し共感を示している応答も散見されたが、偶然であると考えられる。

提案手法に関しては、「比較的意味のある回答が多い」、「会話として成立している」、「あまり違和感がない」、「基本的に相槌を打ち共感を感じた」というポジティブな意見が多かった。対話ログを見ると、ユーザ発話に対して共感を示している応答が多かった。しかし対話ログの中には、共感を示さない応答を行なっている対話も見られた。

共感度が低かった要因を 2 点挙げる。1 つは、ユーザ発話の極性と同一極性のツイートがシステム応答として選択されたが、ユーザの発話内容に対して異なる内容の応答を行って文脈が合わなくなった点である。表 6 のニュース B の提案手法における

共感度の低かった対話例のように、「労災なんてどこでもあるよね」という Negative に判定されたユーザ発話に対し、「「着ぐるみ!?!」そんなわけないよ!」と Negative に判定されたツイートを選択し応答しているが、ユーザの労災の話に対して、着ぐるみに言及した応答を行っており文脈が合っていない。また、ニュース C の提案手法における共感度の低かった対話例のように、「お金がないからね」という Negative に判定されたユーザ発話に対し、「わかるわー。外出面倒だもんね」と Negative に判定されたツイートを選択し応答をしているが、外出をしたくない両者の理由がそれぞれ異なっているため、文脈が合っていない。この問題を解決するために、ユーザ発話に対しツイート群の中から文脈の合うツイートを出力する仕組みを構築し、共感度の向上を目指す。例えば、ユーザ発話から名詞を抽出し、その名詞や類義語を含むツイートを応答することで、文脈の合う対話を行うことが考えられる。

もう 1 つは、ユーザ発話が Neutral と判定された場合、ユー

ザ発話に対して相槌を応答することに共感を感じなかった被験者がいた点である。ニュース B の提案手法における共感度の低かった対話例では、「着ぐるみ大変そう」という Neutral に判定されたユーザ発話に対して「そうだね」と相槌を打っているが、被験者は相槌に共感を感じなかったと思われる。この問題を解決するために、ユーザ発話が Neutral と判定された場合、相槌以外の応答も行い、共感度の向上を目指す。例えば、「私もそう思います。」といった同意を示す応答や、ユーザ発話の一部分を繰り返す応答 [4] を行うことが考えられる。

考察をまとめると、相槌手法やランダム手法では「意味を感じられない」、「応答が一方的に感じた」というネガティブな意見が多く、ログを分析した所文脈の合わない対話が多かった。提案手法では、「会話として成立している」、「あまり違和感がない」というポジティブな意見が多く、ユーザ発話に対して共感を示している応答が多かった。一方で、共感を示さない応答を行なっている対話も見られ、ユーザ発話とシステム応答の文脈が合っていない対話や相槌に共感を感じなかった被験者がいたことが分かった。

6. 結 論

本稿では、ユーザ発話の極性に応じて、ニュース等のコンテンツに関連するツイートを選択して応答する対話システムを提案した。具体的には、ユーザ発話が Positive または Negative であれば、同じ極性のツイートを応答に利用し、ユーザ発話が Neutral であれば、相槌を応答する手法である。

本研究の貢献は以下の 2 点に集約される。1 つは、ウェブニュース記事に関連するツイートから対話を生成しアンケートを用いて評価を行なった結果、ユーザ発話が Positive 若しくは Negative の場合、3 章で挙げた仮説を立証することができたが、ユーザ発話が Neutral の場合は、仮説を立証することができなかった点である。もう 1 つは、複数のニュース記事において、相槌手法、ランダム手法、提案手法の 3 手法を用いて被験者実験を行った所、提案手法は相槌手法及びランダム手法と比較して、いずれも有意に共感度が高いことが確認された点である。

今後の課題として以下の 4 つが挙げられる。第一に、違和感のない自然な対話を実現することである。文脈が合わない不自然な応答を避けるために、ユーザ発話に対し、ツイート群の中から文脈の合うツイートを出力する仕組みを構築する必要がある。第二に、ユーザ発話が Neutral と判定された場合に、相槌以外の多様な応答を含んだ対話を実現することである。ユーザ発話に対して相槌を応答することに共感を感じなかった被験者がいたため、相槌以外の多様な応答を含んだ対話を行う仕組みを構築する必要がある。第三に、システム応答の一貫性を持たせることである。例えば、ニュース記事に関連するツイートの中には、文末表現が敬語であったり平常語であったりと様々なツイートが存在するが、ユーザ発話の文末表現と同じ表現が含まれるツイートを応答するとユーザの満足度が向上する可能性があるため、今後評価を行う予定である。第四に、Negative な対話が続いた際にシステムが話題を転換することである。Negative な対話が続くと、ユーザの気分が落ち込む可能

性があるため、Positive な方向に話題を転換する仕組みを構築する必要がある。

文 献

- [1] りんな <https://www.rinna.jp/>
- [2] Jesse Dodge et al. "Evaluating prerequisite qualities for learning end-to-end dialog systems." arXiv preprint arXiv:1511.06931 (2015).
- [3] 山口貴史, et al. "傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成." 人工知能学会論文誌 31.4 (2016): C-G31.1.
- [4] 石田真也, et al. "傾聴対話システムのための多様な聞き手応答の生成". 情報処理学会第 78 回全国大会 (2016).
- [5] Jiwei Li, et al. "Deep reinforcement learning for dialogue generation." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1192-1202 (2016)
- [6] Hao Zhou, et al. "Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory" Proceedings of the 2018 Conference on AAAI (2018)
- [7] Cloud Natural Language API <https://cloud.google.com/>
- [8] 池田和史, et al. "キュレーションサービスのための主観的コメントの要約手法." DEIM Forum. No. B5-3. 2014.
- [9] 玉瀬耕治, 大塚弥生, and 大谷卓治. "マイクロカウンセリングにおける感情の反映." 奈良教育大学教育研究所紀要 26 (1990): 55-66.
- [10] 水野淳太, et al. "ウェブニュースを利用した雑談対話システム." 人工知能学会言語・音声理解と対話処理研究会資料 55 (2009): 1-6.