

ゲーミフィケーションを利用した効率的な対話ログ収集の試み

叶内 晨 小町 守

首都大学東京

{kanouchi-shin at ed., komachi at}tmu.ac.jp

あらまし 非タスク指向型対話システムにおいて、大規模な対話データを利用した研究・開発が盛んに行われている。しかし、日本語における誰でも利用可能な大規模な対話コーパスは存在しておらず、各々がツイッターなどのマイクロブログやウェブテキストからデータを収集し研究をおこなっている現状がある。そこで本研究では、ゲーミフィケーションを利用することで、大規模な対話ログの収集を目指す。ゲームにおいてユーザ同士の相互刺激を促すことで、ユーザのアクティブ率上昇を狙い、できるだけ自発的にデータが集まる環境を構築する。また発話が破綻しているかを判別する破綻ラベルの自動獲得もおこなうことで、リッチな対話コーパスの作成を目指す。

キーワード 対話システム, ゲーミフィケーション, アノテーション

Toward the effective collection of dialogues using gamification

Shin KANOUCHI and Mamoru KOMACHI

Tokyo Metropolitan University

Abstract In non-task-oriented dialogue systems, there have been thriving researches and developments using large-scale dialogue data. However, there is no large-scale Japanese dialogue data anybody can use, and so each researcher collects a data from microblogs such as Twitter and web texts by him/herself. In this paper, we aim to collect large-scale dialogue data by using gamification. We build an environment where we can collect data as spontaneously as possible. We intend to increase active users by encouraging mutual interaction among users in the game. We plan to create a rich dialogue corpus by automatically acquiring dialogue breakdown labels.

Key words dialogue system, gamification, annotation

1. はじめに

近年、大規模なデータ収集が可能になると共に、データドリブンな対話システムの研究・開発が盛んに行われている。非タスク指向型対話システムにおいては、Twitter を中心とした研究 [1], [2] や大規模な映画の字幕データを利用した研究がある [3]~[5]。これらは大規模な対話データを利用し、多数の発話候補をあらかじめ用意することで、状況に応じて適切に回答するための研究である。しかし、日本語における共通して利用可能な大規模な対話コーパスは存在しておらず、各研究者がそれぞれ Twitter などのマイクロブログやウェブテキストを収集し研究をおこなっている。また、教師データの無いツイートを利用した対話システムを定量的に評価するのは難しいため、対話が成立しているかどうかを示す破綻ラベルを含む大規模な対話コーパスが必要とされる。

一方、パン田一郎 [6], りんな [7] や Tay^(注1) など、SNS を中心

としたチャットボットブームが起きている。それに加え、LINE Messaging API^(注2) や Slack Real Time Messaging API^(注3), Facebook Messenger Platform^(注4) などの API の公開により、対話型のプラットフォームの利用が容易となった。これにより自然言語処理や機械学習に直接関わりのない人たちもチャットボットに興味を持ちつつある。

そこで本研究では、ゲーミフィケーションを利用することで、日本語における (1) 大規模で (2) 公開可能な (3) タグ付きコーパスの開発を目指す。自らチャットボットを作成してみたい人をターゲットにおき、容易にチャットボットを作成することのできるプラットフォームを提供する。その際、ゲーム要素を追加することにより、収集可能なログデータから、効率的に雑談対話コーパスの作成を目指す。さらに、発話文に対して応答文が成り立っているか破綻しているかを判別する破綻ラベ

(注1) : <https://twitter.com/tayandyou>

(注2) : <https://developers.line.me>

(注3) : <https://api.slack.com/rtm>

(注4) : <https://developers.facebook.com/docs/messenger-platform>

ルの自動獲得を可能とし、リッチな対話コーパスを作成する。ゲームにおいてユーザ同士の相互刺激を促しユーザのアクティブ率上昇を狙い、できるだけ自発的にデータが集まる環境を構築する。また、自ら作成したチャットボットを API として外部から呼び出せるようにすることで、対話をするインセンティブを高める。

2. 関連研究

2.1 日本語対話コーパス

日本語における大規模な対話コーパスには、NTCIR Short Text Conversation 日本語タスクのツイートデータがあり、2014 年のツイートペア約 100 万件が扱われている。しかしこれらはツイート ID のみ公開されており、現在クロールを行うと、50 万ツイートペア程度しか抽出できず再現性がない。またツイートにはスパムや誤字・脱字、文法誤りを含むノイズとなる文も数多く含まれている [8], [9]。提案手法ではゲームにおいてユーザが発話にラベルをつけることにより、ノイズが取り除かれることが期待できる。

日本語における中規模な対話コーパスには、Project Next NLP 対話タスクで収集された雑談対話コーパスがある [10]。雑談対話コーパスは人と対話システムにおける約 2 万ペアのコーパスで、対話破綻のアノテーションが付与されている。しかし対話生成において 2 万件は少なく、さらに大きい対話コーパスが必要である。雑談対話コーパスは特定少数のアノテータがタグ付けするコーパスなので、大規模化することが困難であるが、提案手法はゲーミフィケーションを用いてタグ付けするため、サービスの拡大に伴った大規模化が可能である。

本研究では、人と対話システムの対話ペアに対話破綻アノテーションを付与することを目指すため、コーパスの種類は雑談対話コーパスに類似するが、ゲーミフィケーションを利用することでユーザはゲーム感覚で自動的にコーパスが構築されることが特長である。

2.2 ゲーミフィケーション

Higashinaka らは、ユーザ参加型の対話ペア収集の手法として、なりきり質問応答を提案した [11]。ユーザが有名人などの特定の人物になりきりながら質問応答をおこなうことのできるウェブサイト構築することで、好きな人物になりきる楽しさと、好きな人物に対して質問できる面白さから、質問応答を効率的に収集している。

対話ペアの収集以外にも、自然言語処理分野でゲーミフィケーションが使用されている研究がいくつかある。例えば、Kumaran ら [12] は、複数人によるオンラインお絵かきゲームにより言い換えのフレーズを得ることに成功した。Vannella ら [13] は知識ベースを拡充する目的でビデオゲームを作成した。さらに彼らは、ビデオゲームによるアノテーションがクラウドソーシングによるアノテーションよりも質が良いことを証明した。Venhuizen ら [14] は語義の曖昧性を持った単語を、ゲームを利用し複数人と複数選択肢の質問でどう獲得すればよいか示した。

2.3 クラウドソーシング

クラウドソーシングを用いて対話コーパスを構築する研究がある [15]~[17]。Inaba ら [16] は、ある発話文に対してどの応答文が正しいかをクラウドソーシングを利用して選択する際に、選択結果からアノテータの対話力を診断するゲーム機能を導入することで、データの作成と品質管理を同時に行った。本研究は Inaba らと類似するが、ゲーム性が異なるのと、Inaba らはデータの公開を前提としていない。

塚原ら [17] は人間同士の対話において、ワーカーに対話入力と同時にアノテーション付けと校正作業を同時に行う仕組みを提案した。しかし、発話文 1 件あたりの作成時間に平均 2 分ほど要し約 2 ヶ月をかけて 2 万発話を作成しており、大規模なデータを作成するためには金銭的なコストが必要となる。本研究は、ゲーミフィケーションを利用することにより、ユーザはゲーム感覚で自動的にコーパスが構築されることが特徴である。これによりコストを抑えつつ、素早く大規模なコーパスの構築を目指す。

3. ゲーミフィケーションによる対話ログ収集システムの設計と実装

近年、パン田一郎やりんなを中心としたチャットボットブームにより、自然言語処理や機械学習に関わりのない人たちもチャットボットに興味を持ちつつある。一方 1. 節で述べたように、日本語における対話の大規模なデータは公開されていない。

そこで本研究では、自らチャットボットを作成してみたい人をターゲットに、容易にチャットボットを作成することのできるプラットフォームを提供する。その際、ゲーミフィケーションを利用しユーザにチャットボットの育成を行ってもらうことで、収集可能なログデータから対話破綻ラベル付きの雑談対話コーパスの作成を目指す。

3.1 節でゲームの流れと目的について、3.2 節で発話文の生成について、3.3 節で応答文の生成について、3.4 節でユーザに学習させるデータについて、3.5 節でユーザ同士による相互刺激を狙った対戦と評価機能について説明する。

3.1 ゲームの目的と仕様

提案システムで設計するゲームの目的は、自分専用のチャットボットを育成することである。各ユーザは、ゲームに沿いながら与えられた発話文に対する応答文を選択していくことで、自分のチャットボットの応答の幅を増やしていく。

スマートフォンでのプレイを想定し、タップだけで効率よくゲームが成り立つように、システム側であらかじめ応答文の選択肢を用意した。このとき、3.4 節で説明する破綻ラベルを収集する。また、3.5 節で説明する「対戦と評価」機能により、自分のチャットボットが他ユーザのチャットボットと比較してどの程度うまく応答できるかを検証することができる。これによりゲームにおいてユーザ同士の相互刺激を促しユーザのアクティブ率上昇を狙い、できるだけ自発的にデータが集まる環境を構築する。また、自ら作成したチャットボットを API として外部から呼び出せるようにすることで、対話をするインセンティブを高める。

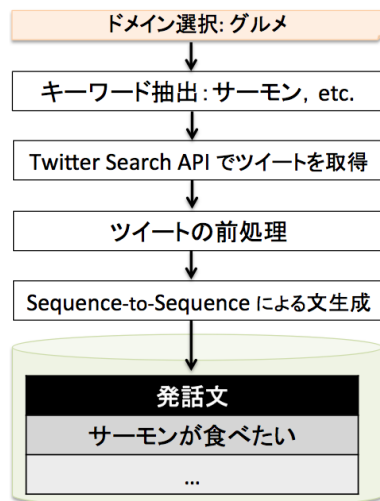


図 1 発話文生成の流れ

表 1 グルメドメインにおいて収集したキーワードの例

| | 採用したキーワード例 | 除外したキーワード例 |
|---|---|--------------------------|
| あ | アーモンド, アイス, アサリ, アジ, あずき, アスパラ, アセロラ, アップル | あざらし, アカシア, アンデス, 阿蘇, 阿波 |
| か | かき氷, かき揚げ, カシス, カステラ, 牡蠣, 海鮮, 蟹, 貝柱, 柿, 鰹, 鰹節 | カジュアル, カプリ, カリフォルニア, 肝臓 |

3.2 発話文の生成

一般的な雑談対話で使用される発話の種類には、挨拶などの日常的なものからユーザの趣味や仕事などの専門的な発話まであり幅が広い。本システムは「自分のチャットボットに発話を学習させて育成する」ことを目的にしたゲームであるが、全てのドメインの発話を学習させるのはユーザにとってコストが大きい。そこで対話のドメインをシステム側で絞ることで、ユーザに対して短期的な目標を設定すると共に少量の学習でもゲームの成果が出やすい仕様にした。

発話文生成の流れを図 1 に示す。発話文を収集するためにあらかじめ雑談のドメインを限定し、そのドメインに応じた適切なキーワードを収集する (3.2.1 節)。収集したキーワードを元にキーワードを含むツイートを収集する (3.2.2 節)。収集したツイートに対して稲葉ら [8] の前処理を行うことで、その発話だけで意味・意図が理解できる文を抽出する (3.2.3 節)。これは、ゲームの設計上、文脈を考慮しなくてよい発話文を収集したかったためである。抽出された文を入力として Sequence-to-Sequence モデル [18] を適応し、バラエティの富んだ発話文を生成する (3.2.4 節)。また、生成を行うことにより、文の著作権の問題に配慮した。

3.2.1 キーワードの選択

今回は実験的にドメインをグルメとした。グルナビにおけるグルメキーワード一覧 (注5) から 38,035 語のキーワードを収集した。収集したキーワードを Google Web 日本語 N グラム 第 1 版 (注6) における 1gram の出現頻度が 10 万回以上の単語 1,277

(注5) : <http://gnavi.joy.ne.jp/g-kw/>

(注6) : <http://www.gsk.or.jp/catalog/GSK2007-C/>

表 2 ルールによって抽出された発話文の例

| キーワード | 発話文の例 |
|-------|---|
| アーモンド | アーモンドが好きです あたまたにアーモンド乗せてる アーモンドってスゴいらしい 間食にアーモンドを食べると頭が冴える |
| かき氷 | 韓国のかき氷初めて食べました かき氷に梅って合うん ナスとかき氷は地球を救うんだ かき氷めっちゃ食べた |

語に限定し、その後、グルメと直接関係がないキーワードを手取り取り取り除き 757 語とした。キーワード例を表 1 に示す。以下の規則によりキーワードを除外した。

- 国名・地名・店名と思われる単語 (e.g. アンデス, 阿蘇)
- 語義の曖昧性が高い単語 (e.g. きつね, キモ)
- 単体ではドメインに直結しない単語 (e.g. カジュアル)
- 2 文字以下のひらがなで構成される単語 (e.g. かつ, もち)

3.2.2 ツイート検索

Twitter Search API (注7) を利用し、キーワードを含むツイートを収集した。今回は 1 キーワード毎に 2,000 件収集した。

3.2.3 ツイートの前処理

ツイートにはスパムや誤字・脱字、文法誤りを含むノイズとなる文も数多く含まれている [8], [9]。また今回はゲームの設定上文脈を考慮しない文を収集するために、稲葉ら [8] を参考にしてツイートの前処理を行った。ツイートを句点および記号で文分割後、各文に対して以下の条件で除外した。ただし単語数や品詞情報を必要とする際には、MeCab (IPADIC) (注8) で分かち書きをした結果を利用した。

- キーワードを含まない文
- 単語数が 4 語未満もしくは 20 語以上の文
- ユーザ名「@」を含む文
- 品詞情報に人名・代名詞を含む文
- キーワードの前後どちらかの単語の品詞が名詞の文 (注9)
- 特定の品詞が不適切な位置に存在する文 (注10)

これらの規則により抽出した文の例を表 2 に示す。キーワードあたり 2,000 件のツイートを収集しているが、前処理を終えた段階で収集できる発話文は、キーワードあたり平均 60 件である。発話文はいくらでも収集し直せるため、規則により Recall よりも Precision に重きを置いて発話文を収集している。

3.2.4 Sequence-to-Sequence モデルによる文生成

Tweet の発話文に対する著作権に配慮するため、Sequence-to-Sequence モデル [18] による生成を行った。前処理をしたツ

(注7) : <https://dev.twitter.com/rest/public/search>

(注8) : <http://taku910.github.io/mecab/>

(注9) : これはキーワードが複合名詞の一部の場合、その複合名詞がキーワードと全く別の意味を表す場合があるため、あらかじめ削除する。(e.g. くるみ 割り人形, 団子 鼻)

(注10) : 記号によりツイートを分割したことによる文の途中で不適切な分割が起こっているものを除外する目的と、打ち間違え・誤字を除外するための規則である。(e.g. 終助詞が文中にある場合)

表 3 Sequence-to-Sequence モデルにより生成された発話文

| 入力文 | 生成した発話文 |
|------------------|---|
| 鮎が届きました | 鮎が届きました 釜めし が届きました 南瓜 が届きました |
| 太刀魚と鰯の漬けがマジで美味しい | 太刀魚と鰯の 煮汁 がマジで美味しい 太刀魚と鰯の 漬けがマジで美味しい ネギ と鰯の 煮汁 がマジで美味しい |
| 鰯のひらきにしよう | 鰯のひらきにしよう 青菜 のひらきにしよう オクラ のひらきにしよう |

イート文に対して Sequence-to-Sequence モデルを適用し、入力文をそのまま出力するように学習した。デコード時にビームサーチを行った後に n ベストをそのまま出力することで、バラエティに富んだ発話文を生成した。本研究では、 $n = 3$ とした。

表 3 に入力文と、それにより生成された発話文を生成確率が高い順に 3 つ示す。1 つ目の例では「鮎が届きました」を入力とした場合、「鮎」が下線で示す他の単語に置き換わっていることがわかる。2 つ目の例は、生成文の第一候補が入力文と異なる例である。本実験では学習に使用されていない文を入力に与えているため、このような現象が起きている。3 つ目の例は、単語を置き換えることで文が不自然になっている。今後の課題として、ルールや共起頻度によりこれらの不自然な文を削除したい。

3.3 応答文の生成

本研究では、DoCoMo の雑談対話 API ^(注11) を利用することで多数の応答文候補をあらかじめ用意した。一つの発話文に対して最低 4 回リクエストを送り、4 件の応答文を得た ^(注12)。これはゲームの設定上、複数の応答文候補が必要となるためである。

応答文をあらかじめ用意するのは、スマートフォンでも快適にゲームが行えること想定し、タップだけで効率よくゲームが成り立つように設計したためである。また、1 つの発言に対して 4 件の応答文を用意することで、応答文を 1 つだけ用意するのと比較して 4 倍のログが収集可能となった。これにより、データ量に対するユーザ側の作業コストを削減している。

またユーザはゲーム内において、雑談対話 API の応答文を選択する以外に自分自身で応答文を作成する機能を実装した。これにより雑談対話 API の応答候補に適切な応答文が存在しない場合にはユーザが自分自身で応答文を作成することで、ある発話文に対して最低一件は適切な応答文を収集可能である。

DoCoMo 雑談対話 API によって生成される文の著作権は自由なため、本研究では DoCoMo 雑談対話 API によって生成される文を応答文としてそのまま利用する。

3.4 チャットボットの学習と破綻ラベル

各ユーザはゲーム内において自分のチャットボットを育成するために、チャットボットに回答の正誤を教える作業を繰り返す。



図 2 チャットボットの学習

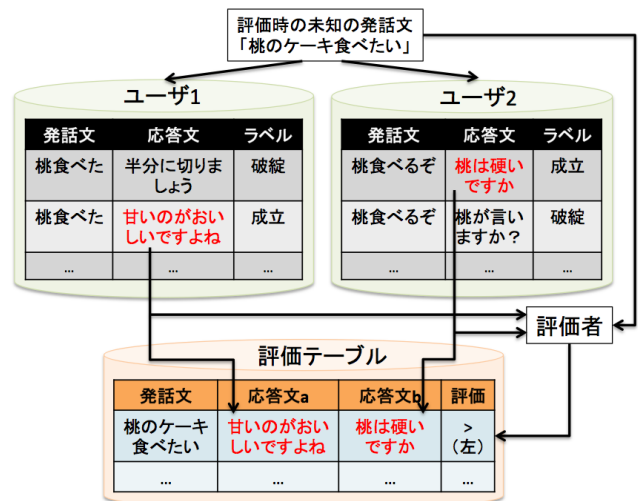


図 3 対戦と評価

図 2 に示すように、任意の発話文とそれに対する複数の応答文が提示されるので、ユーザは自分のチャットボットに回答してもらいたい文を任意の数だけ選択する。これによりある発話文に対する複数の応答文に対して、その応答が正しいかどうかを表す破綻ラベルを収集することができる。学習データはユーザ毎に蓄積し、10 回発話文が投げられるのを 1 セットとする。

発話文に対する複数の応答文は以下の 2 つの方法により選択されている。(1) 発話文に対して DoCoMo の雑談対話 API によりあらかじめ用意された複数の応答文からランダムに選択。(2) 自分が過去に学習させた発話文と応答文のデータベースの発話文の中から、今回の発話文と編集距離が一定以上近い応答文からランダムに選択。

3.5 ユーザ同士による相互刺激

ユーザ同士による相互刺激を促すためにユーザ同士の対戦・評価とランキング機能を実装した。対戦の勝敗によりユーザの

(注11) : <https://dev.smt.docomo.ne.jp>

(注12) : 応答文が重複していた場合は、最大 6 回リクエストを送った。

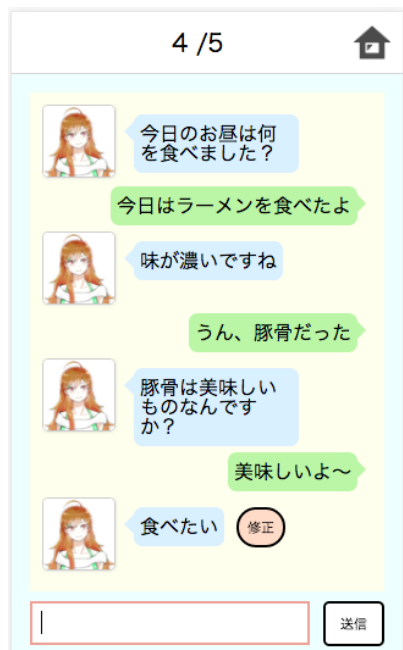


図 4 チャット形式の雑談対話

チャットボットにスコアを与え、ユーザがランキングによって自分のチャットボットが他人と比べてどの程度優れているか相対的に確認できるようにした。これによって、ユーザ間での競争に期待した。また、ゲームにおいて他人のチャットボットを評価することにインセンティブを持たせるために、自分が第三者として対戦を評価した数だけ自分のチャットボットが対戦できるような仕組みとした。

図 3 に対戦と評価の流れを示す。ある発話文が与えられたとき、ユーザ 1 とユーザ 2 のチャットボットはそれぞれ自分の学習データから編集距離が最小の応答文 a・応答文 b を選択する。評価者である第三者は、発話文に対する応答文として応答文 a・応答文 b のどちらがより適切か、「> (大なり)」、「= (イコール)」、「< (小なり)」の 3 択から選択する。この操作を 10 回繰り返し、より適切だと判断された応答文の数が多かったユーザのチャットボットがより優れていたとする。

評価によって得られるログは学習において得られるログとは違い、二つの応答文が与えられたときの相対的な比較である。よってこのゲーミフィケーションの利用により、発話文に対する応答文が絶対的に良いか悪いかのデータだけではなく、他の発話と比べて相対的に良いか悪いかのデータの両方を収集することが可能である。

3.6 チャット形式の雑談の機能

本節では、ある発話文に対して応答文を答えるだけでなく、より現実的な長期的な対話のログ収集を目指す機能について説明する。図 4 に示すように、チャット形式で自分のチャットボットと対話をしながら学習をさせる機能の実装を行った。チャットボットの応答文が正しくないと感じたときに修正ボタンを押すことで、その応答文が正しくなかったことを学習させると共に、どう応答すべきだったのかを入力することで学習させることができる。

4. システムの評価

現時点では被験者による実験をしていないため統計的なデータは得られていないが、収集可能なデータの特徴と、さらに収集できそうなデータについて述べる。

4.1 収集可能なラベル付きデータについて

ゲームによって収集可能なラベル付きデータは 3 種類あり、3.4 節のチャットボットの学習からは (1)「文脈を考慮せず、発話文に対する応答文が絶対的に正しいかどうかのラベル」、3.5 節のチャットボットの評価からは (2)「文脈を考慮せず、発話文に対する応答文が相対的に正しいかどうかのラベル」、3.6 節のチャット形式の雑談からは (3)「文脈を考慮した、発話文に対する応答文が絶対的に正しいかどうかのラベル」が手に入る。

- (1) ユーザがチャットボットを学習させるたびに「文脈を考慮せず、発話文に対する応答文が絶対的に正しいかどうかのラベル」を得る。絶対評価であるため、ユーザが発話文に対する応答文が正しいと思ったかどうかでラベルが決まるが、そのユーザにより破綻の基準やチャットボットのキャラクター性が異なるため、ユーザ毎の解析が必要だと考えられる。また、今回は応答文生成のために DoCoMo 雑談対話 API を利用しているが、応答文の質が不安定で、全ての応答文候補が似たような間違えをしている場合がある。
- (2) 2 人の (1) のデータに対して第三者が評価することにより「文脈を考慮せず、発話文に対する応答文が相対的に正しいかどうかのラベル」を得る。相対評価であるため、評価者によるラベルの揺れは小さいと考えられる。評価者が第三者であるため、できるだけ一般的な応答文を選択した方が相対的に良くなってしまいう可能性が考えられ、独特な応答をするチャットボットが不利になる可能性がある。
- (3) 本研究において、「文脈を考慮した、発話文に対する応答文が絶対的に正しいかどうかのラベル」をうまく生かすための評価を用意することができなかった。また、このデータは、ユーザは文脈を考慮して発話をするにも関わらず、今回のチャットボットは 1 つ前の発話文しか記憶していないため、逆にノイズの多い学習データが増える結果となってしまった。今後の課題として、この機能を活かすための評価を実装するのと、2 件以上前の発話文を応答文生成を行う際に活かす実装にしたい。

4.2 システムの限界

本節では、本システムを最大限活用できた際の限界について議論する。

まず 1 つ目に、発話文の限界の問題がある。現在はドメインを決定後キーワードを収集し、そのキーワードによってツイートをクロール後、規則によって文脈を考慮しない文を抽出している。本システムではゲームの成立を前提としているので、応答文の幅が広がるように、キーワードを含む 4-20 単語の文を抽出してきている。しかし実際の対話では、文脈を考慮しない文においても、今回の条件では起こりえない文が入力になり得

る。よって現実の文と今回の生成した発話文の乖離について調査する必要がある。

2つ目に文脈を考慮しない問題がある。現状における「学習」のシステムは文脈を考慮しておらず、一問一答の形になっている。これは評価のしやすさを重視したためであるが、実際の対話では文脈を考慮しなければならない場合があるため、現実の対話とは乖離がある。しかし、データの収集の仕方とアプリの実装次第では、2件前の自分の発話文と1件前の相手の発話文とそれに対する応答文の3つ組みが正しいかを学習させるシステムが作成可能である。ただし文脈を考慮すればするほど、データがスパースになるため、応答文選択のアルゴリズムにおける工夫が必要となる。現状の2つ組のデータをうまく収集後に、3つ組みも検討したい。

本システムのデータ数における限界は、キーワードに対してツイートをクロールしてくれば良いため、基本的には存在しない。しかしゲーム性を考えた際に、ドメイン毎に適度な達成感を持たせつつ、キーワードとツイートを収集しやすいドメインをあらかじめ選択する必要がある。

4.3 著作権について

本研究では、データの公開を目標としているため、著作権に配慮した上でシステムを設計した。発話文においては、ツイートのデータを前処理により加工後、Sequence-to-Sequenceモデルにより生成することによって配慮した。応答文は著作権がフリーであるDoCoMoの雑談対話APIの文を利用することで対応した。ユーザによる応答文の生成とラベル付けのデータは、ゲームの利用規約として研究目的でのデータの利用をあらかじめ承認してもらうことで解決をした。

5. おわりに

日本語における誰でも利用可能な大規模な対話コーパスは存在しておらず、各々がTwitterなどのマイクロブログやウェブテキストからデータを収集し研究をおこなっている現状がある。そこで本研究では、ゲーミフィケーションを利用することで、低コストで大規模な対話ログの収集を目指す。ゲームにおいてユーザ同士の相互刺激を促すことで、ユーザのモチベーション向上を狙うと同時に、相対的な対話の評価ログを得た。また発話が破綻しているかを判別する破綻ラベルの自動獲得もおこなうことで、リッチな対話コーパスの作成を目指した。

今後の課題として、発話文・応答文生成の質の向上を行った上で、被験者を募り実験的なデータの収集と調整を行う。その後ゲームを公開し、大規模なデータの収集・公開をしたい。また、より現実的な対話の設定に近づけるために、文脈を考慮した対話を収集可能なゲームを設計したい。

文 献

- [1] A. Ritter, C. Cherry, and B. Dolan, “Unsupervised modeling of twitter conversations,” Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp.172–180, 2010.
- [2] R. Higashinaka, N. Kawamae, K. Sadamitsu, Y. Minami, T. Meguro, K. Dohsaka, and H. Inagaki, “Building a conversa-

tional model from two-tweets,” Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on, pp.330–335, 2011.

- [3] R.E. Banchs, “Movie-dic: A movie dialogue corpus for research and development,” Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, pp.203–207, 2012.
- [4] I.V. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau, “Building end-to-end dialogue systems using generative hierarchical neural network models,” Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16), pp.3776–3783, 2016.
- [5] C. Danescu-Niculescu-Mizil and L. Lee, “Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs,” Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (ACL), pp.76–87, 2011.
- [6] 大杉直也, 板澤一樹, 福田基輔, 塩澤繁, 伊豆原大也, “大規模対話システムの一実装と対話ログ傾向分析,” 言語・音声理解と対話処理研究会, pp.67–70, 2014.
- [7] X. Wu, K. Ito, K. Iida, K. Tsuboi, and M. Klyen, “りんな: 女子高生人工知能,” 言語処理学会第22回年次大会発表論文集, pp.306–309, 2016.
- [8] 稲葉通将, 神園彩香, 高橋健一, “Twitterを用いた非タスク指向型対話システムのための発話候補文獲得,” 人工知能学会論文誌, vol.29, no.1, pp.21–31, 2014.
- [9] R. Higashinaka, N. Kobayashi, T. Hirano, C. Miyazaki, T. Meguro, T. Makino, and Y. Matsuo, “Syntactic filtering and content-based retrieval of twitter sentences for the generation of system utterances in dialogue systems,” Situated Dialog in Speech-Based Human-Computer Interaction, pp.15–26, 2016.
- [10] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, 水上雅博, “テキストチャットを用いた雑談対話コーパスの構築と対話破綻の分析,” 自然言語処理, vol.23, no.1, pp.59–86, 2016.
- [11] R. Higashinaka, K. Dohsaka, and H. Isozaki, “Using role play for collecting question-answer pairs for dialogue agents,” INTERSPEECH, pp.1097–1100, 2013.
- [12] A. Kumaran, M. Densmore, and S. Kumar, “Online gaming for crowd-sourcing phrase-equivalents,” Proceedings of the 25th International Conference on Computational Linguistics (COLING), pp.1238–1247, 2014.
- [13] D. Vannella, D. Jurgens, D. Scarfini, D. Toscani, and R. Navigli, “Validating and extending semantic knowledge bases using video games with a purpose,” Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), pp.1294–1304, 2014.
- [14] N. Venhuizen, V. Basile, K. Evang, and J. Bos, “Gamification for word sense labeling,” Proceedings of the 10th International Conference on Computational Semantics (IWCS-2013), pp.397–403, 2013.
- [15] M. Mitchell, D. Bohus, and E. Kamar, “Crowdsourcing language generation templates for dialogue systems,” Proceedings of the INLG and SIGDIAL 2014 Joint Session, pp.16–24, 2014.
- [16] M. Inaba, N. Iwata, F. Toriumi, T. Hirayama, Y. Enoki-bori, K. Takahashi, and K. Mase, “Constructing a non-task-oriented dialogue agent using statistical response method and gamification,” Proceedings of the 6th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART, pp.14–21, 2014.
- [17] 塚原裕史, 内海 慶, “オープンプラットフォームとクラウドソーシングを活用した対話コーパス構築方法,” 言語処理学会第21回年次大会発表論文集, pp.147–150, 2015.
- [18] I. Sutskever, O. Vinyals, and Q.V. Le, “Sequence to sequence learning with neural networks,” Advances in neural information processing systems, pp.3104–3112, 2014.