

# テキストチャットにおけるクエリ終了タイミングの検出

## End-of-Query Detection in Text Chat System

田中 駿  
Shun Tanaka

株式会社サイバーエージェント  
CyberAgent, Inc.

I propose a prediction model for end-of-query detection. It is a common practice that a chatbot system responds an answer when they get a single utterance from the user. A problem arises when a human breaks their query into multiple utterances and a chatbot system can not respond appropriately, leading to lower user experience. In this paper, I propose a prediction method based on LSTM for end-of-query detection.

### 1. はじめに

近年、コールセンターにおける人件費削減などを目的として、顧客からの問い合わせに自動応答を行う対話システムを提供する事業者が増えている。事業者が提供する多くの対話システムでは、ユーザから送られた発話<sup>\*1</sup>に対して適切な返答を用例から検索、もしくは生成し応答する、一問一答形式の対話システムが採用されている。その一方で、人間の発話には1つの会話のまとまり(クエリ)を複数発話に分けて送信する現象が見られる場合があり、複数発話に分割されたクエリに対して、一問一答形式の対話システムでは適切な応答を返せない場合がある。

表1にブログサービスのコールセンターにおける、オペレータとユーザの対話例を挙げる<sup>\*2</sup>。ユーザからの発話、#2、#3は「以前記事を投稿していたブログに記事を投稿しようとしたところ、ログインができなかった(故にログインID、パスワード等の再設定を行いたい)」という趣旨の一つのクエリであるが、ユーザはこのクエリを2つの発話に分割して発話している。理想的には、#2、#3を合わせて一つのクエリと捉え、「ログインできない場合のアカウントの再設定方法」などの回答を提示することが適切と考えられるが、現状の対話システムでは、#2の発話を受信した時点で、用例との類似度検索を行い、「ブログへの記事投稿の方法」などの回答を提示してしまう可能性がある。このように複数発話に分割されたクエリを適切に扱えないことは、応答精度の低下に伴うユーザ満足度(ユーザエクスペリエンス)の低下や顧客流出を招く要因になると考えられる。

そこで本研究では、対話システムがより適切に応答できるよう、ユーザから送信された発話に対して、その発話でクエリとして完結するか否かを予測する手法を提案する。

### 2. 先行研究

Lalaら[Lala 17]は人と音声対話を行うロボットが適切なタイミングであいづちを打てるようにするため、ロジスティック

連絡先: 株式会社サイバーエージェントアドテク本部 AI Lab,  
tanaka.shun\_xa@cyberagent.co.jp

\*1 本稿では、テキストチャットにおいて送信ボタンを押した際に送信される文字列の単位を発話、ユーザとボット・オペレータの一連の発話の集合を対話と呼んでいる。

\*2 情報保護の観点から、実際の対話ログをもとに類似する対話を作成した。

表 1: ユーザとオペレータの対話例

#	発話者	発話内容
1	オペレータ	ログインできないとのことですが、現在どのような状態でしょうか？ ID・登録メールアドレスのアカウント情報と、ログインできない詳細をお教えください。
2	ユーザ	以前に記事を投稿していましたが
3	ユーザ	久しぶりに投稿しようとしたらログインできませんでした。

回帰を用いて直近 500ms 以内にあいづちをすべきかどうかを予測するモデルを提案した。CRF を用いて特徴的な単語に対してフォーカスする手法、決定木を用いてフォーカスした単語に対して質問を返すべきか、単語を復唱するべきかを決定する手法、ロジスティック回帰を用いたロボットが返答すべきタイミングかどうかを判別する手法を提案した。それぞれの手法に対する個別の評価と、それらを併せてアンドロイド Erica に導入した際の評価を行った。

Skantze ら [Skantze 17] は、音声対話において、音声データから素性(音声区間、ピッチ、パワー、品詞など)を抽出し LSTM の入力とすることで連続的に発話の交代を予測する3つのモデルを提案した。タスク指向型の音声対話データである HCRC MapTaskCorpus [Thompson 93] を使用し、発話が止まった際に次の話者を予測するタスクと、発話開始時に一連の発話の長さを予測するタスクで実験・評価を行い、従来のルールベースなどの手法と比べ精度が向上することを示した。

これらの研究は、系列情報をモデル化できる CRF や LSTM などを用いて、ユーザの音声発話データに含まれる情報をもとに発話の転換点を推定するというものであり、言語情報だけでなく、ピッチやパワーなどのいわゆる non-verbal な情報を多く用いて解析を行っている。その一方でテキストチャットにおいては、会話相手の表情や音声など、non-verbal な情報を取得することができず、verbal な情報のみで推定を行う必要があり、言語情報のみを対象にして発話の転換点を推定するタスクは今までに行われていない。

表 2: データセットの内訳

データ セット	対話数	ユーザ 発話数	ラベル 1 の割合	ラベル 0 の割合
学習	22,160	140,170	0.79	0.21
開発	2,770	17,476	0.80	0.20
評価	2,770	17,808	0.78	0.22

表 3: 実験結果

手法	Accuracy	F 値
L2LR	0.778	0.875
LSTM	<b>0.786</b>	<b>0.878</b>

### 3. データ作成

本研究では、弊社の子会社でチャットボット事業を展開している、株式会社 AI メッセンジャー (AIM)\*3 での対話ログを用いてデータセット作成を行った。AIM の提供するサービスでは、ユーザから発話が送信された際、基本的にはボットが一問一答形式での応対を行うが、ユーザの問題を解決できなかった場合や、ユーザがオペレータとの会話を希望した際は人間のオペレータが応対を行うことができるサービスとなっている。

ユーザとオペレータの会話は人間同士の会話であるため、対話相手の発話にまだ続きがありそうだと判断した際は対話相手の発話の続きを待つという行動が見られる。そこで、ユーザとオペレータの応答ログをもとに、特定の発話に続きがあるか否かの 2 値ラベル (end-of-query ラベル) を付与しデータを作成する。end-of-query ラベルは、ユーザからの発話に続きがある場合に 0、発話に続きがなくオペレータに話者が交代している場合に 1 を付与した\*4。

2016 年 9 月 25 日から 2017 年 9 月 24 日までの 1 年分のデータを収集した結果、計 27,700 対話のデータが得られた。それらのデータから学習、開発、評価データセットを作成するために対話数が 8:1:1 となるようにランダムにデータを分割しデータセットの作成を行った。データセットの内訳を 2、ラベル付けされた対話データの例を 5 に示す。

### 4. 予測モデル

本研究ではベースラインとしてロジスティック回帰 (L2LR) を用い、LSTM を用いた手法との比較を行う。それぞれの手法の概要を以下に示す。

- L2LR :  
単語 unigram(表層形) に対するベクトルを素性としロジスティック回帰を学習したもの
- LSTM :  
単語 unigram(表層形) に対して分散表現ベクトルを求め、LSTM の入力としたもの

なお、LSTM は対話におけるユーザが発した直前の発話情報は保持せず、各発話毎に end-of-query の予測を行うものとした。

### 5. 実験

#### 5.1 実験設定

形態素解析器として MeCab[Kudo 05]、形態素解析用辞書として mecab-ipadic-neologd[Sato 15] を用いた。ロジスティック回帰の学習には LIBLINEAR[Fan 08] を使用した。提案手

\*3 <https://www.ai-messenger.jp/>

\*4 適切なクエリのまとまりが取れない場合もあるが、アノテーションコストと比較した結果、自動構築できる方策を取った。

表 4: L2LR において重みの多くかかった単語例

単語	重み
何	0.01797
言わ	0.01202
か	0.01017
また	0.007
ページ	0.00592
まし	0.0059
ログイン	0.00484
ん	0.0046
ませ	0.00449
しよ	0.00298

法における単語分散表現は日本語 Wikipedia のダンプデータから予め word2vec[Mikolov 13] を学習したものを使用した。提案手法のモデル学習には Adam[Kingma 14] を用いた。なお、モデルの学習の際、単語の分散表現は学習対象外とした。モデルの隠れ層の次元  $n = \{32, 64, 128\}$  でそれぞれ実験し、単語分散表現の次元を 300、イテレーション数  $i = 150$  を上限として実験を行った。

#### 5.2 評価方法

手法の評価方法として、予測ラベルと正解ラベルの一致率である Accuracy と、Recall と Precision の調和平均である F 値を用いる。

#### 5.3 パラメータチューニング

L2LR, LSTM とともに、開発データセットを用いて最も高い F 値を出すパラメータを検証した。その結果、L2LR では、正則化パラメータのコスト  $c = 2^{-5}$ 、閾値  $\theta = 0.39$ 、LSTM では、隠れ層の次元  $n = 128$ 、イテレーション数  $i = 146$  となった。

#### 5.4 実験結果と考察

それぞれの手法における評価データセットに対する実験結果を 3 に示す。

ベースラインである L2LR と提案手法である LSTM を比較すると、Accuracy, F 値のどちらにおいても LSTM が L2LR を上回る結果となった。対話例全文に対する各手法による予測結果を表 5、対話例中に出現する 115 単語において L2LR での重みの大きい上位 10 単語を表 4 に示す。L2LR では、“ます”の活用形である“まし”などに重みがかかっていることが分かるが、時系列を考慮できるモデルではないため、#2「以前に記事を投稿していましたが」の発話に対しても end-of-query の確率が高くなっていることが伺える。その一方で、LSTM では単語の系列を考慮することができるため、#2 の発話に対して end-of-query の確率を低く予測することができていることが分かる。

### 6. まとめと今後

本研究では、ユーザとオペレータの対話ログを用いて、テキストチャットにおけるユーザの発話がクエリとして完結してい

るか否かの予測に取り組んだ。実験の結果、ロジスティック回帰を用いたモデルより LSTM を用いた提案手法の精度が高くなり、発話に含まれる単語の系列をモデルに組み込むことでわずかながらも精度が向上することがわかった。

今後はこのモデルを改良し、ユーザ毎の発話特徴やユーザの一連の発話、オペレータの直前の発話内容の情報などを加味したモデルを実現したいと考えている。また、実システムへの導入を行った結果の UX の評価や、雑談対話のデータにおける実験も行いたいと考えている。

## 参考文献

- [Fan 08] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J.: LIBLINEAR: A Library for Large Linear Classification, *Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874 (2008)
- [Kingma 14] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *CoRR*, Vol. abs/1412.6980, (2014)
- [Kudo 05] Kudo, T.: MeCab : Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/> (2005)
- [Lala 17] Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., and Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking, in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 127–136 (2017)
- [Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, Vol. abs/1301.3781, (2013)
- [Sato 15] Sato, T.: Neologism dictionary based on the language resources on the Web for Mecab (2015)
- [Skantze 17] Skantze, G.: Towards a General, Continuous Model of Turn-taking in Spoken Dialogue using LSTM Recurrent Neural Networks, in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 220–230 (2017)
- [Thompson 93] Thompson, H. S., Anderson, A., Bard, E. G., Doherty-Sneddon, G., Newlands, A., and Sotillo, C.: The HCRC Map Task Corpus: Natural Dialogue for Speech Recognition, in *Proceedings of the Workshop on Human Language Technology, HLT '93*, pp. 25–30, Stroudsburg, PA, USA (1993), Association for Computational Linguistics

表 5: オペレータとユーザの対話例

#	発話者	発話内容	end-of-query ラベル	LSTM	L2LR
1	オペレータ	ログインできないとのことですが、現在どのような状態でしょうか？ ID・登録メールアドレスのアカウント情報と、ログインできない詳細をお教えてください。	-	-	-
2	ユーザ	以前に記事を投稿していましたが	0	0.320	0.516
3	ユーザ	久しぶりに投稿しようとしたらログインできませんでした。	1	0.571	0.673
4	オペレータ	アカウント情報をお教えいただけますか？ ID または、ブログ URL ・登録メールアドレス ※現在登録メールアドレスがご利用できない場合は、その旨もお教えてください。	-	-	-
5	ユーザ	ID もパスワードも忘れてしまっていて	0	0.433	0.593
6	ユーザ	ID を忘れたときのページで自分のメールアドレスを入力しても	0	0.554	0.565
7	ユーザ	メールアドレスが違うと言われます。	1	0.676	0.641
8	オペレータ	ログインしたいアカウントを確認しますので、 ・ブログ URL ・ブログタイトルとニックネーム ・その他登録覚えのあるメールアドレス 上記をお教えいただけますか？	-	-	-
9	ユーザ	https:// ID / アカウント名	1	0.643	0.573
10	オペレータ	ご連絡ありがとうございます。ご連絡のアカウントは、ドメイン ( メールドメイン名 ) で登録がございます。登録メールアドレスにお心あたりはございませんか？	-	-	-
11	ユーザ	メールアドレス でしょうか？	1	0.697	0.814
12	オペレータ	一致いたしました。登録メールアドレスの変更を行いますので、ご利用可能なメールアドレスをお教えいただけますか？	-	-	-
13	ユーザ	メールアドレス お願いします。	1	0.905	0.925
14	オペレータ	お待たせしました。登録メールアドレスの変更を行いましたので、パスワード再設定をお願いいたします。パスワード再設定: 再設定 URL	-	-	-
15	ユーザ	ありがとうございます！ 書き込む事が出来ました！	0	0.645	0.888
16	ユーザ	また何かありましたら、宜しくお願いします。本当にありがとうございました。	1	0.896	0.957
17	オペレータ	無事解決のご連絡ありがとうございます！	-	-	-