

相手の発話を深掘りするための質問生成技術

Question Generation to Deepen Your Talk

片山太一*¹ 大塚淳史*¹ 光田航*¹ 齋藤邦子*¹ 富田準二*¹
Taichi KATAYAMA Atsushi OTSUKA Koh MITSUDA Kuniko SAITO Junji TOMITA

*¹日本電信電話株式会社 NTT メディアインテリジェンス研究所
NTT Meddia Intelligence Laboratories, Nippon Telegraph and Telephone Corporation

1. はじめに

スマートフォンによる音声エージェントやコミュニケーションロボットの実用化に伴い、エージェントやロボットと自由な対話をするための対話技術の必要性はますます高まっている。特に近年では、エンターテインメントやカウンセリングを目的とした雑談対話に関する研究 [2][5][1] が増加している。

このような雑談対話において、東中ら [1] や目黒ら [8] では、シナリオ型対話と統計型対話を組み合わせたハイブリット型の雑談対話システムを提案しており、どのような発言をしても、ある程度会話が破たんしない返答ができるシステムを現実的に可能になっていることが報告されている。

上記の関連研究では、表層的につながる対話を目指して、「話題を深掘りできること」に主眼をおいた研究はほとんどない。相手の発話を深掘りするためには、相手の発話に対して、さらなる情報を引き出すための質問をする必要がある。実際に目黒ら [9] では、聞き役に徹した場合と雑談をした場合の対話行為の違いを分析しており、聞き役を担当者が、雑談をしていたときと比較して、質問をした回数が大幅に増えていることを報告している。

質問方法の代表的な分類に、オープン・クエスチョンとクローズド・クエスチョンと言うものがある。相手が「はい、いいえ」または「A か B か」の択一で答えられるような、回答範囲を限定した質問の仕方をクローズドクエスチョンといい、これに対し、「どう思うか?」などのように、制約を設けず相手に自由に答えさせるような質問のしかたをオープン・クエスチョンという。一般的に、相手の考えや事実を明確にしたいときは、聞き返し等のクローズド・クエスチョンが有効で、相手からより多くの情報を引き出したい場合では、オープン・クエスチョンが有効であると言われている。その為、話題を深掘りするための質問として、オープン・クエスチョンが有効と考えられる。オープン・クエスチョンの質問を作成する場合、原則として、5W1H (いつ/どこ/だれ/なに/なぜ/どうやって) で質問を作成する。

我々は、5W1H の疑問詞に着目し、相手の発話に対して、適切な疑問詞を用いて質問を生成することで、相手の発話を深掘りすることができる質問文の生成が可能であると考えた。

そこで、本論文では、twitter のリプライの関係性と疑問詞に着目し、相手の発話に対して、5w1h の質問をしているデータを抽出し、学習データを自動的に作成することで、適切な 5w1h の疑問詞を利用した相手の発話に対する深掘りをする質

問文をディープラーニングによる手法を用いて生成できることを示す。

本実験では、相手の発話に対して正しい深掘りの質問が生成されているのと同時に、深掘りするための質問のバリエーションについても広く生成できていることを示す。

1.1 関連研究

5W1H を用いた質問文を生成する際に第一に考えられるのは、Why の疑問詞を使った質問文を生成することである。その他の疑問詞は、相手の発話によって使用してよい場合とよくない場合がある。例えば、「この間、横須賀に旅行に行った」と相手が言っているのに対して、「どこに行きましたか」のような場所を聞く質問や「いつ行きましたか」のような日付を聞くような質問はできないので、where や when といった疑問詞が使用できないのがわかる。それに対して、why を用いた「何の目的で横須賀に行ったのですか」は相手の深掘りの質問として妥当である。Why のなぜやなんでといった疑問詞のみは相手が発話した内容がなぜ起きたのか、なんでそのような事になっているのかということを知る質問文を生成できるので、相手の発話がどんな発話であっても利用することができる。

しかし、毎回なぜやなんでを質問文として生成してしまうと、相手の情報を深掘りをする質問はできているが、毎回同じ質問のみになってしまう。東中ら [7] では、対話が破綻しているかどうかのアノテーションを手で行っており、同じ内容の発話をしているものは対話が破綻したとみなしている。そのため、相手の発話に対して深掘りの質問を生成する際には、「なぜ」や「なんで」といった質問のみではなく、様々な疑問詞を利用した質問を生成する必要がある。

疑問詞を用いた質問文を生成するタスクとして、タスク対話の発話生成が挙げられる [4]。タスク対話においては、すでに相手に聞くべきスロットが決まっており、相手の発話を理解し、埋まっていないスロットがわかるため、埋まっていないスロットを聞くための質問文を適切な疑問詞を用いて生成することで適切な質問を生成することができる。しかし、今回の想定する対話は雑談対話であり、雑談対話では、相手の発話に対して聞くべきスロットが決まっておらず、どういった質問が適切なかが自明ではない。

その為、雑談における相手の発話を深掘りするための質問文の生成には、どういった疑問詞を使い、どういった項目についての質問をすればよいのかを推定する必要がある。上記の問題を解くために、我々は、どういった疑問詞を使い、どういった項目についての質問をすればよいのかを同時に推定してくれる有用な技術であるディープラーニングベースの手法 [3] を用いる。しかし、ディープラーニングベースの手法では、学習データが重要であり、相手の発話を深掘りするための質問生成のタ

連絡先: 片山太一, 日本電信電話株式会社 NTT メディアインテリジェンス研究所, 〒 239-0847 神奈川県横須賀市光の丘 1-1, katayama.taichi@lab.ntt.co.jp

スクに合わせた学習データが世の中に存在しておらず、学習データを用意することが非常にコストがかかるといった問題が挙げられる。

2. 提案手法

本節では、2.1 で、質問文の生成に必要な学習データの自動獲得について述べ、2.2 でディープラーニングによる質問生成の手法について述べる。

2.1 twitter からの質問文の獲得

今回作成したいデータは相手の発話と相手の発話を深掘りをするための質問文である。上記のようなデータを自動で獲得する際に、東中ら [6] でも使用されている twitter データのリプライ関係を利用する。はじめに、相手の発話、相手の発話に対する質問、質問に回答した発話の 3 つ組のデータを取得した。相手の発話と相手の発話に対する質問の 2 つ組でなく、3 つ組を取得した理由は、相手の発話に対する質問である 2 つ目の発話で終了している対話、つまり相手の発話に対する質問に回答していない対話は相手の深掘りをしていないと想定されるため対象外とし、質問に回答している 3 つ組のデータを取得した。質問文の生成に必要な学習データは、相手の発話と相手の発話に対する質問であるため、上記で取得した 3 つ組のデータから 1 つ目のデータを相手の発話文、2 つ目のデータを相手の発話に対する質問文として抽出し、学習データとした。

具体的には、相手の発話、相手の発話に対する質問、質問に回答した発話の 3 つ組のデータは、twitter のリプライ関係を利用し、ユーザ A → ユーザ B → ユーザ A といったリプライが 2 回続いている関係性を持つ 3 つ組かつ、ユーザ B の発話が 5w1h で質問しているものを相手の発話、相手の発話に対する質問、質問に回答した発話の 3 つ組のデータとして抽出した。ユーザ B の発話が 5w1h で質問しているものかどうかの判定については、「いつ、どこで、なにが、どうして」等の 5w1h の疑問詞が出現するかクエスチョンマークで終了するものといった条件で抽出した。

2.2 ニューラルネットに基づく質問文の生成

本論文での質問文の生成には、vinyals ら [3] で述べられている Encoder-Decoder によるニューラル翻訳モデルを応用したニューラルネットに基づく発話生成技術を利用する。機械翻訳における入力言語と出力言語の関係性を相手の発話と相手の発話に対する質問とし、相手の発話に対する質問文を生成した。

3. 評価実験

3.1 データ

2.1 節の手法を用いて、2016 年の 5ヶ月分の twitter のデータから 2 つ目のデータに 5w1h の疑問詞が存在する 3 つ組発話を取得すると、530 万件のデータが抽出でき、そのうちの 500 万件を利用した。この 3 つ組のうち、1 つ目のデータを入力言語側の学習データ、2 つ目のデータを出力言語側の学習データとして使用した。上記で作成したデータを 100 万、200 万、500 万のデータ・セットに分け、それぞれを利用して評価を行う。

学習に必要なバリエーションデータとして、東中らが収集した雑談コーパス [1] を利用する。このコーパスには、初対面の話者が 1 対 1 でテキストチャット形式の雑談を行った対話が収録されている。この収録されたデータをもとに、2.1 節の手法と同様に、5w1h の疑問詞が出現するかクエスチョンマークで終了する文正規表現を用いて抜き出し、抜き出した文の前の

文を相手の発話、抜き出した文を相手の発話に対する質問として、約 6000 件のデータを抽出した。

テストデータとしては、人手を用いて、自分の経験や事実、属性について友人や知り合い等に話かける内容のテキストを 100 文作成した。

3.2 ベースライン

すべての入力に対して、「なんで？」を質問として返却するシステムをベースラインとする。

3.3 評価方法

本論文では、評価方法は 2 種類の観点で評価をする。一つ目は、生成された質問文が入力したテストデータに対する質問として妥当かどうかを判断する指標値である。妥当性の評価は人手による主観評価を行った。

$$\text{妥当性} = \frac{\text{妥当と判定された質問文の数}}{\text{テストデータの数}} \quad (1)$$

二つ目は生成された質問文がどのくらいバリエーション広く生成できたのかを判断する指標値である。

$$\text{ユニーク率} = \frac{\text{質問文の表記のバリエーションの数}}{\text{テストデータの数}} \quad (2)$$

上記 2 つの指標で生成した質問文が妥当かつバリエーションが広く生成されているのかを測定する。

4. 評価結果と考察

生成された質問文が入力したデータに対する質問として妥当かどうかの結果とどのくらい重複したかの結果を表 2 に示す。また、実際生成された質問文の例を表 1 に示す。2 の結果から、ベースラインの手法である「なんで」を質問文とした場合の妥当性は非常に高いが、ユニーク率が非常に低い値となっている。提案手法では、学習量が多くなるに連れて、妥当性とユニーク率が高い数値になっていることがわかる。これは、バリエーション広く、有用な質問が生成できていることを示している。学習データの量が増えると妥当性とユニーク率が上がっているが、「なんで」が出現する頻度は実は増えている。学習データの量が多いと「なんで」が出てくるようなものを人手で確認した所、学習データが少ないときには、間違った質問をしているものが、学習データの量が増えた所で「なんで」に変わっていることがわかった。学習データが少ないときには、入力文と似ている文が悪い影響を与えて、間違った質問をするが、学習データの量が多くなることで、曖昧なものは「なんで」になるような学習されていると予想される。また、表 1 の「お正月に太りました」の例のように、学習データが少ないと、「なんで」といった疑問詞のみを用いて質問が生成されているものは、学習データが多くなると「何キロ？」といった具体的な質問になるような例が多く見られた。図 1 に、それぞれの学習データのときの妥当だった質問の数と妥当かつ疑問詞のみの数を示す。学習量が増えることで、妥当な質問が増える。また、疑問詞があまり数として差がないということは、学習量の増加により、疑問詞のみではなく、より具体的な質問が生成できていることを示している。

5. おわりに

本論文では、雑談対話の相手の発話を深掘りをするための質問文の生成について取り組んだ。twitter のリプライ関係と

表 1: 出力結果

| 入力文 | 100万 | 200万 | 500万 |
|--------------|------|--------|------------|
| 趣味は読書です | なんの？ | なんの？ | 何読むの？ |
| フクロウカフェに行きたい | どこ？ | どこ？ | どこにあるの？ |
| メガネをかけています。 | どこ？ | どんな感じ？ | どんなメガネですか？ |
| お正月に太りました。 | なんで？ | 何キロ？ | 何キロ？ |

表 2: 評価の結果

| データの数 | 妥当性 | ユニーク率 |
|--------|------|-------|
| ベースライン | 1.0 | 0.01 |
| 100万 | 0.58 | 0.4 |
| 200万 | 0.63 | 0.39 |
| 500万 | 0.76 | 0.54 |

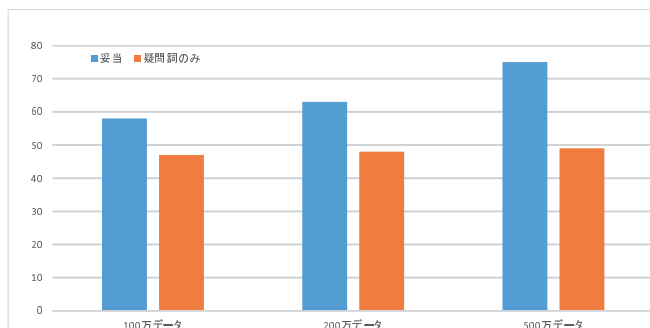


図 1: 妥当数と妥当な質問文のうちの疑問詞のみの数

疑問詞に着目して自動作成した学習データを利用することで、正しくかつバリエーション豊かな深掘りをする質問文を生成する技術を提案した。評価実験の結果、数百万文程度の学習データを用意することで、正しくかつバリエーション豊かな質問文を生成できることを明らかにした。

現在は、twitter のデータしか学習データに使用していないため、「なんで」、「いつ」、「どこで」といった疑問詞のみを答えるような質問文を生成してしまっている。評価実験のバリエーションデータに使用した東中らが収集した雑談コーパス [1] の中の質問文を手で確認した所、「どんな音楽が好きですか?」「何か食べたいものがありますか」等の疑問詞+文で作成されているものが多く、これは、実際の対話では、疑問詞のみではなく、ある程度の長さを持った質問をする方が自然であることを示している。

今後は、twitter のような短いデータのみではなく、長い発話も学習データに加えることで、ある程度長い質問も生成することで、バリエーションを更に広げていくことに取り組む予定である。また、質問文生成の精度及びバリエーションの向上と共に、1問1答形式のみでなく、複数発話の情報、またそれ以前にユーザの情報として構造化して持っているものを利用し、相手の情報を聞き出していく発話を生成する技術についても取り組む予定である。

参考文献

- [1] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. Towards an Open Domain Conversational System Fully Based on Natural Language Processing. *Proc of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 928–939, 2014.
- [2] Alan Ritter, Colin Cherry, and William B. Dolan. Data-driven response generation in social media. *Proc of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 583–593, 2011.
- [3] Oriol Vinyals and Quoc Le. A neural conversational model. *Proc of ICML Deep Learning Workshop 2015*, 2015.
- [4] Jason Williams and Steve Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech and Language*, 21:393–422, 2007.
- [5] Wilson Wong, Lawrence Cavedon, John Thangarajah, and Lin Padgham. Mixed-initiative conversational system using question-answer pairs mined from the web. *Proc of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)*, pages 2707–2709, 2012.
- [6] 東中竜一郎, 川前徳章, 貞光九月, 南泰浩, 目黒豊美, 堂坂浩二, and 稲垣博人. 2 ツイートを用いた対話モデルの構築. 言語処理学会年次大会, 2012.
- [7] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, and 水上雅博. 雑談対話データの収集と対話破綻アノテーション及びその分類. 自然言語処理, 2015.
- [8] 目黒 豊美, 杉山弘晃, 東中竜一郎, and 南泰浩. ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築. 第 28 回人工知能学会全国大会論文集 (JSAI 2014), 2014.
- [9] 目黒 豊美, 東中竜一郎, 堂坂浩二, 南泰浩, and 磯崎秀樹. 聞き役対話システムの構築を目的とした聞き役対話の分析. 情報処理学会自然言語処理研究会 (NL), 2009.