

# ソーシャルメディア上のインタラクションを利用した オープンドメイン対話応答生成

Ritvik Choudhary 河原大輔  
早稲田大学理工学術院  
{ritvik@fuji.,dkw@}waseda.jp

## 概要

人間のような豊かな会話能力を持つオープンドメイン対話システムの構築は、自然言語生成における課題の一つである。しかし、近年のこの分野の発展にもかかわらず、既存のオープンドメイン対話生成モデルは知識を取り込むことができず、未知の発話に対して反復的または典型的な応答をすることが多い。知識に基づく対話応答生成に関する先行研究は、主にペルソナの組み込みや、Wikipedia のような構造化された知識ベースの検索に焦点を当てている。本研究は、より汎用的かつ単純なアプローチをとり、SNS に見られる社会的なインタラクションを通じて人間の応答行動を模倣することで、システムの雑談能力を向上させることを目的とする。提案手法では、Reddit から抽出されたコメントデータを大規模な知識源とみなし、ニューラル retriever を用いて関連文を抽出する。続いて、その関連文を付加文脈として発話文に連結し seq2seq generator に渡す。このような generator と retriever からなるモデルをオープンドメイン対話データセットを用いて学習し、自動及び人手評価した結果、提案手法の有効性を確認した。

## 1 はじめに

Siri、Google Assistant などの音声アシスタントの生活への浸透に伴い、雑談対話システムの重要性が増している。そのため、雑談対話における応答生成は、自然言語処理における主要なタスクの一つと言える。

しかしながら、既存の雑談対話応答生成モデルは、面白くない、あるいは有益な応答が得られないといった基本的な問題に依然として悩まされている [1, 2]。この主な原因は、人間とは異なり、モデルは個性や知識などを持たないためと考えられる。

この問題に対処するために、近年の研究では、ペルソナ [3, 4]、知識ベース [5] などの付加的な文脈情報を考慮することが提案されている。これらに対して、本研究は、応答生成モデルの雑談能力を向上させるという、より一般的な立場からこの問題に取り組む。特に、人間が会話する実際の環境、すなわち、社会的インタラクションを利用することを考える。そのために、ソーシャルメディア上の多様な会話からなる外部知識ベース SMikb (Social Media Interactions knowledge base) を構築する。SMikb は、既存の知識ベースと異なり、自然言語で記述されており、応答生成において外部知識として利用する。

提案モデルは、retriever と generator からなる。retriever は、入力された発話で SMikb を検索し、関連知識を得る。generator は seq2seq モデルからなり、入力発話の情報に加えて関連知識も踏まえて応答を生成する。具体的には、Dense Passage Retriever [6] (以下、DPR) を retriever に、BART [7] を generator に用い、これらを同時に学習する。オープンドメインの対話データセットで generator と retriever を学習し、Reddit の投稿とコメントからなる SMikb を retriever の知識源として利用する。

応答生成の実験を行ったところ、提案手法が既存の seq2seq ベースライン (BART) を自動および人手評価指標のすべてにおいて上回ることがわかった。SMikb を利用することで、生成された応答はより魅力的になるだけでなく、より適切かつ自然になることも示され、提案手法の有効性が確認された。

## 2 関連研究

**生成モデル** 近年、Transformer [8] や BERT [9] に始まるブレイクスルーにより、自然言語処理研究は急速に変化した。これらに自己回帰 seq2seq モデル (T5 [10]、BART) が続き、対話システムなどの生成モデルの性能が大幅に向上した。本研究では BART

を強力なベースラインとして採用する。

**知識に基づく応答生成モデル** 対話システムに文脈や外部知識などを組み込むことが、近年、大いに注目されている。PersonaChat [4] や EmpatheticDialogues [11] では、ペルソナや感情が考慮されている。また、知識ベースを利用した研究も進んでいる [2, 5]。対話以外では、外部知識を様々な QA タスクで利用する研究がある [12, 13]。応答生成のために検索を行うという点で、本研究に最も近いのは、retriever を別途学習し、学習データから検索した応答候補も generator に入力し生成するアプローチ [14] である。本研究は、retriever 単独の学習を必要とせず、かつ、生成時により大規模な外部知識ベースを検索するという点で大きく異なる。また、Wikipedia などに基づく retriever を用いた知識対話生成においては、事実との矛盾を調査した研究もなされている [15]。本研究では、応答の内容が事実 に即しているかというよりは、対話システムの雑談能力の向上に焦点を当てる。

### 3 提案手法

本研究では、より自然で人間らしい応答生成のため、ソーシャルメディア上のインタラクションを外部知識ベース (SMIkb) として導入する。提案モデルは SMIkb を利用した retriever-generator の 2 段階構成である。

#### 3.1 タスク設定

外部知識に基づく応答生成のタスクは、入力発話  $\mathbf{u}$  と関連知識を含む可能性のある文書集合  $\mathcal{D}$  が与えられたとき、 $m$  単語の応答  $\mathbf{r} = (r_1, r_2, \dots, r_m)$  を予測するタスクとして定義することができる。本研究では、入力発話  $\mathbf{u}$  と知識ベース  $\mathcal{D}$  に対して、モデルが確率  $p(r_i | \mathbf{r}_{<i}, \mathcal{D}; \theta)$  で応答  $\mathbf{r}$  を生成するように学習する ( $\theta$  はモデルパラメータを表す)。

#### 3.2 モデル

先行研究 [12, 16] における retriever に基づく QA システムにならい、本研究の応答生成タスクにおいても retriever と generator の 2 段階構造を採用する。図 1 に提案モデルの概要を示す。図に示した通り、本モデルの対話応答生成は、2 段階のプロセスとして定式化する。(1) retriever はあらかじめ索引付けされた SMIkb から入力発話に関連する上位  $k$  文書を検索し、(2) generator は検索された文書と発話が連結

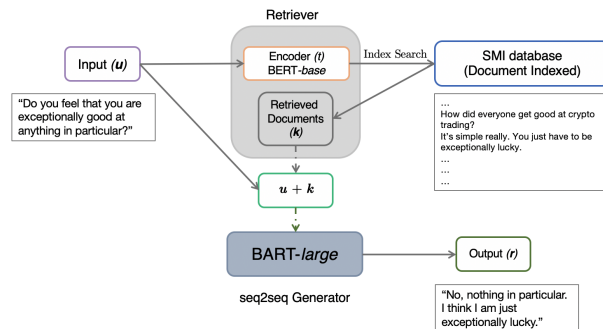


図 1: 提案モデルの概要

されたものが入力され、応答を生成する。

具体的には、事前学習済みの Dense Passage Retriever (DPR) を retriever とし、大規模コーパスで事前学習済みの seq2seq モデル BART を generator として利用する。DPR は入力クエリ (発話)  $\mathbf{u}$  に対する上位  $k$  個の関連文書を検索するように学習された高効率なニューラル retriever である。この効率化は、クエリと知識ベース全体を独立した 2 個の BERT [9] のエンコーダーに通すことによって実現されている (図の  $t$ )。FAISS [17] ライブラリを用いて SMIkb のインデックスとなる埋め込みベクトルを事前に構築する。事前に構築することで、オフラインでの高速な検索が可能になる。

3.1 節のタスク設定に従い、発話  $\mathbf{u}$  と知識ベース  $\mathcal{D}$  からの上位  $k$  文書  $d_j$  が与えられたときに、モデルが応答  $\mathbf{r}$  を生成する確率を次のように定義する。

$$p(\mathbf{r} | \mathbf{u}; \theta, \lambda) = \sum_j^k p_\lambda(d_j | \mathbf{u}; \lambda) \prod_i p_\theta(r_i | \mathbf{u}, \mathbf{r}_{<i}, d_j; \theta), \quad (1)$$

ここで、 $\theta$  と  $\lambda$  は、それぞれ generator と retriever のパラメータを表す。これらは共に end-to-end で学習される。検索される SMIkb においては正解の文書がタグ付けされていないため、関連文書を潜在変数とみなす。従って、デコーディングの際には、ビームサーチを用いて最も確率の高い最良の応答を返すために、検索された全ての文書に対してこの確率を周辺化する。

## 4 実験

提案モデルを、様々な外部知識データセットとオープンメイン対話データセットを用いて比較する。また、純粋な BART による二つのベースラインとの比較も行う。

表 1: 対話データセットの規模

データセット	合計ターン数	Train	Valid	Test
DailyDialog	76,743	53,721	11,511	11,511
DailyDialog++	39,913	27,939	5,987	5,987
Cornell Movie-Dialogs	221,088	154,762	33,163	33,163
Reddit (疑似対話データ)	200,000	140,000	30,000	30,000

## 4.1 SMikb

本研究では、対話システムの雑談能力を向上させることを目的として、人間の会話行動を蓄積した知識ベース SMikb を構築する。SMikb は、アメリカのニュース掲示板サイトである Reddit<sup>1)</sup> の 2020 年 (1 月～11 月) における人気スレッドタイトルとその上位 100 件のコメントから構成する。SMikb を構築するために、オープンソースの Pushshift API [18] を通じ、合計 160 万件のデータをスクレイピングした。SMikb の一部を付録 A に載せる。

さらに、Reddit のような SNS から構築した SMikb の有効性を検証するために、他に 2 種類の知識ベースを構築し比較する。一つは、Wikipedia 記事のタイトルと冒頭 100 語からなる同サイズ (60 万エントリを無作為に抽出) の Wikipedia 知識ベース (以下、「Wiki」) である。もう一つは、上記二つのデータベースの半分ずつを 1:1 で組み合わせたもの (以下、「Mix」) である。

## 4.2 対話データセット

本研究では様々なオープンドメインの対話データセットと Reddit からスクレイピングした対話データセットを用いて、提案モデルの fine-tuning を行う。

**オープンドメイン対話データセット** 高品質な日常対話データセットとして DailyDialog [19] と DailyDialog++ [20]、また、映画の台詞で構成された Cornell Movie-Dialogs Corpus [21] の 3 種類をマージして用いる。

**Reddit** さらに、SMikb とは別に Reddit から 20 万件のコメントペアを抽出し、擬似的な対話データセットとして対話データを補完する。

上記のデータセットはマルチターン対話であるが、本研究の設定にそぐわないため各シングルターンを対話として抽出した。使用するデータセットの規模を表 1 にまとめる。

## 4.3 実験設定

**モデル** 提案する retriever-generator モデルは、事前学習された Dense Passage Retriever と BART-large (24 層、406M パラメータ) からなり、対話データセットを用いて fine-tuning する。モデルの学習には Adam optimizer を用い、バッチサイズ 1、学習率  $3 \times 10^{-5}$  のハイパーパラメータで学習する。retriever の上位  $k$  検索には  $k = 3$  を設定し、生成の戦略としてビームサイズ 5 のビームサーチを用いる。

**ベースライン** 提案モデルの有効性を確認するために、retriever なしの BART-large による二つのベースラインを使用する。一つ目は、4.2 節で述べたデータセットのみで学習し、外部知識データベースの SMikb を用いないものである (以下、「ベースライン 1」)。二つ目として、retriever-generator の 2 段階構造を通して外部知識を利用することの有効性を確認するため、SMikb の全データ (タイトルとそのコメント) を学習データに結合し、この新しい拡張データセットで純粋な BART-large を fine-tuning する (以下、「ベースライン 2」)。

## 4.4 評価

外部知識の影響を確かめるため、生成された対話応答について自動評価と人手評価の両方を行う。それらの結果を表 2 にまとめる。

**自動評価** 関連性の指標である BLEU [22] スコアに加えて、生成された応答の多様性について、Distinct-N [1] も算出する。

**人手評価** 自動評価指標は、生成された対話応答の実際の質とあまり相関がないことが先行研究で報告されている [23]。本研究では、Amazon Mechanical Turk<sup>2)</sup> を用いてクラウドソーシングを行い、生成結果について以下の三つの評価指標で人手評価も行う。

- 関連性 (Relevance): その応答が発話に関連しているかどうか
- エンゲージメント (Engagement): その応答をした相手とさらに会話を継続したいかどうか
- 知識 (Knowledge): その応答が知識に富んでいるかどうか

評価者となるクラウドワーカーには、テストデータから無作為に選んだ 100 個の応答を上記の各指標に

1) Reddit: <https://www.reddit.com>

2) Amazon Mechanical Turk: <https://www.mturk.com>

表 2: 生成された応答の自動評価と人手評価の結果。モデル (テストデータを除く) の中で最もスコアが良いものに「\*」を付与した。ODD は 4.2 節の **Open-Domain Dataset** の集合である。

モデル設定	学習 データ	知識ベース	自動評価			人手評価		
			BLEU-4	Dist-1	Dist-2	Relevance	Engagement	Knowledge
Gold (テストデータ)	-	-	-	-	-	3.50	<b>3.33</b>	<b>3.47</b>
ベースライン 1	ODD	なし	1.31	0.20	0.96	2.82	2.35	3.00
ベースライン 2	ODD + SMikb	なし	1.05	0.12	0.47	3.03	3.02	2.89
提案手法 (SMikb)	ODD	SMikb	<b>9.78</b>	<b>2.80</b>	<b>16.90</b>	3.70	<b>3.18*</b>	3.35
提案手法 (Wiki)	ODD	Wiki	6.93	2.57	14.91	3.74	3.16	3.39
提案手法 (Mix)	ODD	SMikb + Wiki	6.03	2.45	14.08	<b>3.83</b>	3.01	<b>3.41*</b>

表 3: 各モデルにおける応答生成の例

発話	応答 (SMikb)	応答 (Wiki)	応答 (Mix)	応答 (ベースライン)
Here comes the birthday cake.	Wow, what a lovely cake. I'll have a piece.	Wow, what a beautiful cake. It must have tasted as good as I expected.	Really? Wow. Happy birthday to you! You look so healthy.	Thank you.
I'd better go. I gotta get up early. Go over to my parents'.	What are you going to do for breakfast?	You mean you don't wanna go with me?	Why don't you stay at my place? I'll cook dinner.	I don't think so.
Something's burning?	Oh my goodness, it's on fire!	Oh my goodness, it's burning!	Oh, my bread! It's burning!	No, it's not.
Would you mind if we just called it an early night?	Even I couldn't. I have work tomorrow.	I'd love to. But I've got to work tomorrow.	What are you talking about? We just ate?	I don't think so.

ついて 1~5 の尺度で採点するよう求める。各応答は 7 人の評価者によって採点し、最終的なスコアはそれらの平均値とする。

## 4.5 結果

自動評価では、生成時に retriever を通じて SMikb を利用する提案手法が、生成される応答の質を向上させ、より多様な応答を可能にすることが確認された。また、すべての評価指標において 2 種類のベースラインより提案手法の方が高性能であることがわかった。さらに、Reddit ベースの SMikb モデルは他の組み合わせよりも優れており、高品質な応答の生成につながったと考えられる。

クラウドソーシングによる人手評価でも同様の傾向が見られ、提案モデルの全ての組み合わせに対するスコアがベースラインよりも高かった。ブートストラップ法に基づく有意性検定により、全ての観点においてベースライン 2 と提案手法それぞれの人手評価スコアの差を検証したところ、統計的に有意であることが分かった ( $p < 0.01$ )。当初の仮説通り、SMikb モデルはエンゲージメントや雑談能力の面で最も高いスコアを記録した。一方、SMikb と

Wikipedia の組み合わせ (Mix) は、会話の関連性と「知識性」のバランスが取れており、関連性のスコアが人間の応答に対するスコア (Gold) を上回った。これは、本研究の対象であるシングルターン評価の場合、SMikb に Wikipedia を加えた提案手法による応答はクラウドワーカーの評価者にとってより関連性があるように感じられたためと考えられる。

各モデルによる生成応答の例を表 3 に示す。

## 5 おわりに

本研究では、対話システムの雑談能力を向上させることを目的として、SNS 上のインタラクションからなる SMikb を備えた retriever-generator 2 段モデルを提案した。提案手法は、ニューラル retriever と seq2seq 生成モデルを一括で fine-tuning し、生成時には関連知識を知識ベースから抽出することによって、より自然な対話応答を実現する。実験では、本研究で構築した SMikb に基づく応答生成は、自動評価とクラウドソーシングによる人手評価の両方によって、より多様で適切、かつ魅力的であることが示され、ベースラインと比較して性能の改善が確認された。

## 謝辞

本研究は LINE 株式会社との共同研究の助成を受けて行った。

## 参考文献

- [1] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 110–119, San Diego, California, June 2016. Association for Computational Linguistics.
- [2] Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. A knowledge-grounded neural conversation model. **arXiv preprint arXiv:1702.01932**, 2017.
- [3] Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 994–1003, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [4] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2204–2213, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [5] Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. Wizard of wikipedia: Knowledge-powered conversational agents. In **International Conference on Learning Representations**, 2019.
- [6] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6769–6781, Online, November 2020. Association for Computational Linguistics.
- [7] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [10] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2020.
- [11] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 5370–5381, Florence, Italy, July 2019. Association for Computational Linguistics.
- [12] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 9459–9474. Curran Associates, Inc., 2020.
- [13] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading Wikipedia to answer open-domain questions. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1870–1879, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [14] Jason Weston, Emily Dinan, and Alexander Miller. Retrieve and refine: Improved sequence generation models for dialogue. In **Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI**, pp. 87–92, Brussels, Belgium, October 2018. Association for Computational Linguistics.
- [15] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. Retrieval augmentation reduces hallucination in conversation, 2021.
- [16] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. Realm: Retrieval-augmented language model pre-training, 2020.
- [17] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. **arXiv preprint arXiv:1702.08734**, 2017.
- [18] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In Munmun De Choudhury, Rumi Chunara, Aron Culotta, and Brooke Foucault Welles, editors, **Proceedings of the Fourteenth International AAI Conference on Web and Social Media, ICWSM 2020, Held Virtually, Original Venue: Atlanta, Georgia, USA, June 8-11, 2020**, pp. 830–839. AAAI Press, 2020.
- [19] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. DailyDialog: A manually labelled multi-turn dialogue dataset. In **Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 986–995, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.
- [20] Ananya B. Sai, Akash Kumar Mohankumar, Siddhartha Arora, and Mitesh M. Khapra. Improving dialog evaluation with a multi-reference adversarial dataset and large scale pretraining. **Transactions of the Association for Computational Linguistics**, Vol. 8, p. 810–827, Dec 2020.
- [21] Cristian Danescu-Niculescu-Mizil and Lillian Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In **Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics**, pp. 76–87, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- [22] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [23] Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2122–2132, Austin, Texas, November 2016. Association for Computational Linguistics.

表 4: SMikb の一部

SMikb	
タイトル	コメント文
LPT: If you borrow something like a tool or a generator from someone, return it in BETTER shape than you got it.	My dad always said that returning something in the same condition you received it is the absolute bare minimum.
SoftBank Nears \$40 Billion Deal to Sell Arm Holdings to Nvidia	Nvidia is priced decently for what they offer.
Apple to Give Employees Paid Time Off to Vote in U.S. Election	This exactly. A large majority of disenfranchised communities work jobs that don't observe federal holidays.
Apple may be working on a foldable iPhone	I can confirm that Apple would be stupid to not be working on one. Whether they ever release one is up for debate, but they're definitely working on one.
Anyone else feel like there are so many good games that are completely spoiled by a single bad mechanic?	I don't have an issue with all inventory management, I have an issue with limited capacities. I'd much rather be able to pick it all up and sort it later in town. At least mods can fix that.
Google changed my device trade in value from \$350 to \$17.50.	Isn't it crazy we need good Samaritans to step in and help with these things because Google CS can't?

## A SMikb

Reddit に基づく SNS データベース (SMikb) の一部を表 4 に示す。検索はタイトルに対して行うが、応答生成の根拠となる付加文脈として本質が会話と近いコメント文を利用する。

## B 人手評価: MTurk

応答生成の人手評価を行うため、Mturk 上タスク支持率 (過去のタスクによるクラウドワーカーの評価値) が 90%以上の US 在住の英語話者に依頼した。平均報酬は 1 問あたり 1 セントに設定した。