

令和 3 年 5 月 19 日現在

機関番号：32689

研究種目：基盤研究(B) (一般)

研究期間：2017～2020

課題番号：17H01830

研究課題名(和文) ナゲットに基づくタスク指向対話の自動評価に関する研究

研究課題名(英文) Nugget-based Automatic Evaluation of Task-oriented dialogues

研究代表者

酒井 哲也 (Sakai, Tetsuya)

早稲田大学・理工学術院・教授

研究者番号：80723519

交付決定額(研究期間全体)：(直接経費) 11,300,000円

研究成果の概要(和文)：評価型国際会議NTCIR (NII Testbeds and Community for Information access Research)にて対話タスクを運営し、最終成果物としてヘルプデスク対話データセットDCH-2を構築した。これに関連し、WebDB Forum 2018論文賞Runner-upを受賞した。また、対話品質スコアの分布を推定させるタスクのための評価指標を設計し、有効性を示した。その成果を含む論文は、自然言語処理分野のトップ会議ACL-IJCNLP 2021にフルペーパーとして採択された。

研究成果の学術的意義や社会的意義

我々が構築した対話データセットDCH-2は利用許諾書にサインするだけで誰でも利用できる。元データは中国語であるが全てに人手による英訳を付与しているため汎用性が高い。さらに、我々が提案した対話品質スコアの分布を推定するタスクのための評価指標は、ordinal quantificationと呼ばれる形態のすべてのタスク(「ネガティブ」「ニュートラル」「ポジティブ」のような順序尺度上の分布を扱うタスク)に適用可能である。

研究成果の概要(英文)：We served as task organisers of NTCIR (NII Testbeds and Community for Information access Research), an international evaluation forum, and constructed the DCH-2 dialogue data set. Our paper on an early version of this data set received the excellent paper runner-up award at WebDB Forum 2018. We also designed an evaluation measure for a task that requires systems to estimate the gold distribution of dialogue quality scores and showed its effectiveness. The work has been accepted as a full paper at ACL-IJCNLP 2021, a top conference in natural language processing.

研究分野：情報検索、自然言語処理、インタラクション

キーワード：対話 ヘルプデスク ナゲット ordinal quantification データセット

1. 研究開始当初の背景

コールセンターやヘルプデスクの機能を完全に自動化できれば、顧客に対する均質かつ 24 時間稼働のサービスが低コストで実現できる。このためにはヘルプデスク対話により問題解決に至るまでの過程を分析する要素技術や、ヘルプデスクシステムによる同じ発話が異なるユーザにどのように捉えられるかを予測する要素技術構築が不可欠である。しかし、当時は上記のような要素技術の研究に適したデータセットも対話品質を定量化する評価指標も確立されていなかった。

2. 研究の目的

本研究では、多様なユーザの要求に応え問題解決に導くヘルプデスク対話システムの構築に役立てるために、アノテーションつき大規模対話データセットを構築・公開し、同時に、対話中の問題解決につながる発話(ナゲット)をもとに対話品質を定量化する手法を提案・評価する。さらに、多様なユーザの観点から対話品質の推定システムを評価する方法を具体化する。

3. 研究の方法

データセットの構築については、評価型国際会議 NTCIR (NII Testbeds and Community for Information access Research) において対話タスクを運営し、これに向けて中国語・英語パラレルのヘルプデスク対話データを収集し、発話レベルのナゲットアノテーション(どのような発話が問題解決に結びついたかを示す)と、対話レベルの対話品質アノテーションを約 20 名のアノテータより付与した。データのソースは中国語ソーシャルメディアの Weibo であり、各対話の英語版は人手の翻訳により収集した。

構築したデータセットの新規性のひとつに、全アノテータの多数決によりラベルをひとつ決定するのではなく、各アノテータのラベルをそのまま保持することによりラベルの分布を正解データとして定義した点が挙げられる。例えばアノテータが 3 人である場合、3 名がそれぞれ (-1, 0, 1) 点をつけた場合でも、(0, 0, 0) 点をつけた場合でも、多数決ではラベルが 0 となってしまう。一方、本研究では、評価がばらついている場合と、評価が一致している場合を区別することができる。このアプローチは、同じシステム発話に対し異なる評価を下すユーザがいることを前提としてシステムを設計するのに重要であると考えられる。

評価指標の開発については、ナゲットを利用して対話の情報量を定量化するアプローチと、上記対話品質の分布を活用して推定分布との比較を行うアプローチの 2 つを検討した。なお、対話品質のような順序尺度上の分布を推定するタスクは ordinal quantification と呼ばれる。

4. 研究成果

(文献[7]より転載)

Table 1: DCH-1 and DCH-2 dialogue corpus statistics.

	DCH-1 [11]	DCH-2
Data timestamps	Jan. 2013 - Sep. 2016	Jan. 2013 - Sep. 2016 (DCH-1 corpus) Oct. 2016 - Apr. 2018 (NTCIR-14 STC-3 NDDQ test dialogues [9]) Apr. 2018 - Jul. 2019 (NTCIR-15 DialEval-1 test dialogues [10])
#Chinese dialogues	3,700	4,390 (DCH-1 + 390 STC-3 + 300 DialEval-1)
#English translations	1,264 (34%)	4,390 (100%)
#Helpdesk accounts	161	161
Avg. #turns/dialogue	4.162	4.201
Avg. turn length (#chars)	48.31	54.541
# Annotators per dialogue	19	20

(1) 中国語・英語ヘルプデスク対話データセット DCH-2 の構築、ナゲットに基づく対話評価方法の設計

我々はナゲットに基づく評価という着想[1]をもとに、NTCIR-14 Short Text Conversation Task [5] および NTCIR-15 Dialogue Evaluation Task [6] を運営し、これを通じて対話データセットの拡充を進め、最終成果物としてヘルプデスク対話データセット DCH-2 を構築した[7]。これは user agreement form の提出により誰でも利用可能である。上記 Table 1 は、最終成果物である DCH-2 データセットと、その途中経過である DCH-1 の各種統計量を示したものである。

なお、中間報告として国内会議 WebDB Forum 2018 により発表した論文[4] は、論文賞 Runner-up を受賞した。

(2) Ordinal Quantification Task 用の新しい評価指標 R(S)NOD の設計と有効性の検証

各対話の対話品質スコアを約 20 名のアノテータにより収集し、スコアの真の分布を参加チームに推定させる上記 NTCIR タスクの対話品質サブタスクのために、ordinal quantification に適した新たな評価指標 Root (Symmetric) Normalised Order-aware Divergence を設計し[2][3]、その有効性を示した。その成果を含む論文は、自然言語処理分野のトップ会議 ACL-ICJNLP 2021 にフルペーパーとして採択された。Ordinal quantification に適している既存の評価指標 (Earth Mover 's Distance の一種) と比べ、提案指標が統計的に安定していることを実験的に示すことができた。

< 引用文献 >

[1] Zeng, Z., Luo, C., Shang, L., Li, H., and Sakai, T.: Test Collections and Measures for Evaluating Customer-Helpdesk Dialogues, Proceedings of EVIA 2017, pp.1-9, 査読あり, 2017.

[2] Sakai, T.: Towards Automatic Evaluation of Multi-Turn Dialogues: A Task Design that Leverages Inherently Subjective Annotations, Proceedings of EVIA 2017, pp.24-30, 査読あり, 2017.

[3] Sakai, T.: Comparing Two Binned Probability Distributions for Information Access Evaluation, Proceedings of ACM SIGIR 2018, pp.1073-1076, 査読あり, 2018.

[4] Zeng, Z., Luo, C., Shang, L., Li, H., and Sakai, T.: Towards Automatic Evaluation of Customer-Helpdesk Dialogues, Journal of Information Processing, Volume 26, pp.768-778, 査読あり, 2018. **WebDB Forum 2018 論文賞 Runner-up**

[5] Zeng, Z., Kato, S., and Sakai, T.: Overview of the NTCIR-14 Short Text Conversation Task: Dialogue Quality and Nugget Detection Subtasks, Proceedings of NTCIR-14, pp.289-315, 査読なし, 2019.

[6] Zhaohao Zeng, Sosuke Kato, Tetsuya Sakai, and Inho Kang: Overview of the NTCIR-15 Dialogue Evaluation (DialEval-1) Task, Proceedings of NTCIR-15, pp.13-34, 査読なし, 2020.

[7] Zhaohao Zeng and Tetsuya Sakai: DCH-2: A Parallel Customer-Helpdesk Dialogue Corpus with Distributions of Annotators' Labels, arXiv:2104.08755, 査読なし, 2021.

[8] Tetsuya Sakai: Evaluating Evaluation Measures for Ordinal Classification and Ordinal Quantification, Proceedings of ACL-IJCNLP 2021, to appear, 査読あり, 2021.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Zhaohao Zeng, Cheng Luo, Lifeng Shang, Hang Li, Tetsuya Sakai	4. 巻 26
2. 論文標題 Towards Automatic Evaluation of Customer-Helpdesk Dialogues	5. 発行年 2018年
3. 雑誌名 Journal of Information Processing	6. 最初と最後の頁 768-778
掲載論文のDOI（デジタルオブジェクト識別子） 10.2197/ipsj.jip.26.768	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計9件（うち招待講演 0件/うち国際学会 9件）

1. 発表者名 Sosuke Kato and Rikiya Suzuki and Zhaohao Zeng and Tetsuya Sakai
2. 発表標題 SLSTC at the NTCIR-14 STC-3 Dialogue Quality and Nugget Detection Subtasks
3. 学会等名 NTCIR-14（国際学会）
4. 発表年 2019年

1. 発表者名 Zhaohao Zeng and Sosuke Kato and Tetsuya Sakai
2. 発表標題 Overview of the NTCIR-14 Short Text Conversation Task: Dialogue Quality and Nugget Detection Subtasks
3. 学会等名 NTCIR-14（国際学会）
4. 発表年 2019年

1. 発表者名 Tetsuya Sakai
2. 発表標題 Comparing Two BinnedProbability Distributions for Information Access Evaluation
3. 学会等名 ACM SIGIR 2018（国際学会）
4. 発表年 2018年

1. 発表者名 Zeng, Z., Luo, C., Shang, L., Li, H., and Sakai, T.
2. 発表標題 Test Collections and Measures for Evaluating Customer-Helpdesk Dialogues
3. 学会等名 The 8th International Workshop on Evaluating Information Access (EVIA 2017) (国際学会)
4. 発表年 2017年

1. 発表者名 Sakai, T.
2. 発表標題 Towards Automatic Evaluation of Multi-Turn Dialogues: A Task Design that Leverages Inherently Subjective Annotations
3. 学会等名 The 8th International Workshop on Evaluating Information Access (EVIA 2017) (国際学会)
4. 発表年 2017年

1. 発表者名 Zhaohao Zeng, Sosuke Kato, Tetsuya Sakai, and Inho Kang
2. 発表標題 Overview of the NTCIR-15 Dialogue Evaluation (DialEval-1) Task
3. 学会等名 NTCIR-15 (国際学会)
4. 発表年 2020年

1. 発表者名 Ting Cao, Fan Zhang, Haoxiang Shi, Zhaohao Zeng, Sosuke Kato, Tetsuya Sakai, Injae Lee, Kyungduk Kim, and Inho Kang
2. 発表標題 RSLNV at the NTCIR-15 Dialogue Evaluation (DialEval-1) Task
3. 学会等名 NTCIR-15 (国際学会)
4. 発表年 2020年

1. 発表者名 Junjie Wang, Yuxiang Zhang, Tetsuya Sakai, and Hayato Yamana
2. 発表標題 SKYMN at the NTCIR-15 DialEval-1 Task
3. 学会等名 NTCIR-15 (国際学会)
4. 発表年 2020年

1. 発表者名 Tetsuya Sakai
2. 発表標題 Evaluating Evaluation Measures for Ordinal Classification and Ordinal Quantification
3. 学会等名 ACL-IJCNLP 2021 (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

対話データセットDCH-2 Zhaohao Zeng and Tetsuya Sakai: DCH-2: A Parallel Customer-Helpdesk Dialogue Corpus with Distributions of Annotators' Labels, arXiv:2104.08755, 査読なし, 2021.
--

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
中国	清華大学			
韓国	Naver Corporation			