

Overview of Dialog System Technology Challenge 10

吉野 幸一郎^{1,2,a)}

概要: 本講演では、対話システム関連タスクの国際シェアードタスクとして行っている Dialog System Technology Challenge 10 (DSTC10) について述べる。まず DSTC の歴史的経緯について説明し、今回提案されたタスクの詳細や、シェアードタスク・ワークショップの今後の方向性について議論する。

1. はじめに

Dialog system technology challenge (DSTC) は、元々は dialog state tracking challenge という名前で、dialog state tracking (DST; 対話状態推定) のオープンベンチマークに対するシェアードタスクとして始まった [1]。その後、第5回までは DST のシェアードタスクとして、マルチドメインへの対応 [2]、未知ドメインへの適応 [3]、知識ベースへのアクセス [4]、多言語的応 [5] などの課題をもってベンチマークを構築してきた。これらのデータは DST に関するオープンベンチマークとして様々な研究で利用され、その思想は MultiWOZ [6], [7], [8] へと継承された。

一方 DSTC は第6回 [9] からより広範な対話関連タスクを対象とする目的で現在の dialog system technology challenge へと名前を変え、DST だけでなく対話応答生成や破綻検出などを広範にカバーするようになった。また、第6回からの大きな変化はタスクの公募を行うことで様々な企業が参入する形になったことで、これまでに Facebook、三菱、NTT、IBM、Microsoft、Google、Amazon、WeChat などの国際企業がタスク提案と同時にデータセットの作成に参与するようになった。これ以外にもいくつかの大学からのタスクも採択されており、企業と大学が組んでオープンベンチマークを作成するようなケースも見られるようになった。これは対話システム研究のオープンサイエンス化という観点からは非常に重要な変化であり、また対話システム研究がより実世界で役立つ可能性について議論され始めてきた証でもある。これ以降、第7回からは毎年12月から2月頃にタスク公募、5月から9月頃にシェアードタスクの開発・検証フェーズを行い、次の年の国際会議 AAAI でワークショップを行うという流れが定着して [10], [11], [12]

今回で第10回目の開催となる。

多様な企業、大学、研究機関等から様々なタスクが提案されることにより、対話データのオープン化は大きく進んだ。一方で、行われるタスクが企業が推進する特定のプラットフォームやタスク定義に偏ってしまうという問題や、主たる開発フェーズの時期から参加者の偏りが見られるなど、様々な問題も議論されるようになってきている。また、対話研究において真に必要なとされるオープンベンチマークは何かということについても、継続的な議論が必要である。加えて、基本的にこうしたシェアードタスクは英語のデータで行われるため、日本での対話研究・対話システム開発に役立てるためには、どのようにローカライズ・翻訳を行うかという点も重要である。本講演では第10回目で行われたシェアードタスクの概要と、今後求められる対話システム研究のオープンベンチマークについて議論する。

参考文献

- [1] Williams, J. D., Henderson, M., Raux, A., Thomson, B., Black, A. and Ramachandran, D.: The dialog state tracking challenge series, *AI Magazine*, Vol. 35, No. 4, pp. 121–124 (2014).
- [2] Henderson, M., Thomson, B. and Williams, J. D.: The second dialog state tracking challenge, *Proceedings of the 15th annual meeting of the special interest group on discourse and dialogue (SIGDIAL)*, pp. 263–272 (2014).
- [3] Henderson, M., Thomson, B. and Williams, J. D.: The third dialog state tracking challenge, *2014 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, pp. 324–329 (2014).
- [4] Kim, S., Luis Fernando D' Haro, Banchs, R. E., Williams, J. D., Henderson, M.: The fourth dialog state tracking challenge, *Dialogues with Social Robots*, Springer, pp. 435–449 (2017).
- [5] Kim, S., D' Haro, L. F., Banchs, R. E., Williams, J. D., Henderson, M. and Yoshino, K.: The fifth dialog state tracking challenge, *2016 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, pp. 511–517 (2016).
- [6] Budzianowski, P., Wen, T.-H., Tseng, B.-H., Casanueva,

¹ 理化学研究所ガーディアンロボットプロジェクト

² 奈良先端科学技術大学院大学

^{a)} koichiro.yoshino@riken.jp

- I., Ultes, S., Ramadan, O. and Gasic, M.: MultiWOZ-A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 5016–5026 (2018).
- [7] Eric, M., Goel, R., Paul, S., Sethi, A., Agarwal, S., Gao, S., Kumar, A., Goyal, A., Ku, P. and Hakkani-Tur, D.: MultiWOZ 2.1: A Consolidated Multi-Domain Dialogue Dataset with State Corrections and State Tracking Baselines, *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 422–428 (2020).
- [8] Zang, X., Rastogi, A., Sunkara, S., Gupta, R., Zhang, J. and Chen, J.: MultiWOZ 2.2: A Dialogue Dataset with Additional Annotation Corrections and State Tracking Baselines, *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pp. 109–117 (2020).
- [9] Hori, C., Perez, J., Higashinaka, R., Hori, T., Boureau, Y.-L., Inaba, M., Tsunomori, Y., Takahashi, T., Yoshino, K. and Kim, S.: Overview of the sixth dialog system technology challenge: DSTC6, *Computer Speech & Language*, Vol. 55, pp. 1–25 (2019).
- [10] Luis FernandoD’ Haro, Yoshino, K., Hori, C., Marks, T. K., Polymenakos, L., Kummerfeld, J. K., Galley, M., Gao, X.: Overview of the seventh dialog system technology challenge: DSTC7, *Computer Speech & Language*, Vol. 62, p. 101068 (2020).
- [11] Kim, S., Galley, M., Gunasekara, C., Lee, S., Atkinson, A., Peng, B., Schulz, H., Gao, J., Li, J., Adada, M. et al.: The eighth dialog system technology challenge, *arXiv preprint arXiv:1911.06394* (2019).
- [12] Gunasekara, C., Kim, S., D’Haro, L. F., Rastogi, A., Chen, Y.-N., Eric, M., Hedayatnia, B., Gopalakrishnan, K., Liu, Y., Huang, C.-W. et al.: Overview of the ninth dialog system technology challenge: Dstc9, *arXiv preprint arXiv:2011.06486* (2020).