Logic-Guided Neural Utterance Generation from Drone Sensory Data
(Extended Abstract)[⋆]

Stefan Borgwardt[2], Ernie Chang[1], Kathryn Chapman[1], Vera Demberg[1], Alisa Kovtunova[2], and Hui-Syuan Yeh[1]

[1] Department of Computer Science, Saarland Informatics Campus, Germany
`{cychang@coli,s8kachap@teams,demberg@lst,yehhui@coli}.uni-saarland.de`
[2] Chair for Automata Theory, Technische Universität Dresden, Germany
`firstname.lastname@tu-dresden.de`

Drone technology and drone control have recently advanced rapidly, to the point that consumer drones with impressive features and capabilities are commonplace [4]. Advanced sensors and improved control algorithms have made flying drones much simpler and many drone applications have become possible (e.g. aerial surveys, mapping, aerial movies or selfie-drones). As they are used for an increasing range of tasks, interacting with drones becomes more important.

To enable these interactions in everyday life, it is essential to devise a natural language generation (NLG) setup that can flexibly process a variety of data collected by the drone and convey only the important information to the user. In this paper, we propose a neural generation model (or drone assistant) that verbalizes messages from sensor data records in order to perform a controlled handover to a human drone pilot (see Figure 1). Recent data-driven methods have achieved good performance on various NLG tasks [2, 3, 6]. However, most studies focus on surface descriptions of simple record sequences, e.g. attribute-value pairs of fixed or very limited schemas, such as E2E [7] and WikiBio [5]. In contrast, in our setup there is a large variety of data records, and the *content selection task* is substantially harder (only critical information, not all available information, should be mentioned at handover time). Moreover, it is desirable that the system generalizes well to diverse as well as unseen environments during its operation.

To this end, we argue that it is necessary to leverage intermediate content representations to achieve faithful and controllable text generation in different environments. In this paper, these intermediate representations are generated using description logic (DL) ontologies [1]. This removes the burden of logical reasoning from the neural model and allows for more flexible and high quality utterances to be produced.
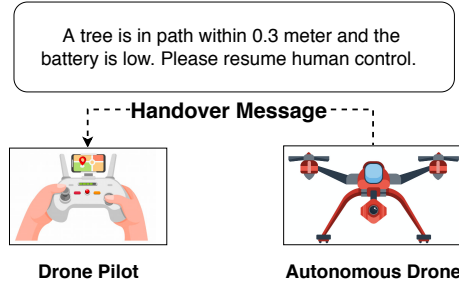
Our Contribution

We study this approach on a new dataset called **DroneParrot** that consists of 316 data records derived from real drone footage across 8 environments, such as

---

Fig. 1. We focus on the drone handover as the main communicative function.

Urban and Ocean. Each sensor data record is combined with a natural language utterance and an intermediate representation generated by DL reasoning. The raw sensor data includes values such as wind speed, altitude, temperature, battery level, and information about nearby objects. To select the critical information for a handover, we define a DL ontology that derives additional information about the current situation. Several high-level concepts like RiskOfPhysicalDamage are defined to identify critical situations, using axioms like

$$\exists near.Object \sqcap \exists environment.LowVisibility \sqsubseteq RiskOfPhysicalDamage.$$

For a data record (i.e. an ABox) that entails RiskOfPhysicalDamage(drone), we compute all ABox justifications of this assertion. We do not include TBox axioms in these justifications since there is no time to explain the whole reasoning process during a handover situation. The union of all obtained justifications is a subset of the original data record, which forms the *intermediate representation* in our dataset. This shorter representation is used by the neural model to generate a focused handover message. The reduced size of the intermediate representation enables learning with fewer training samples and better generalization across different environments.

The drone ontology and an example of two videos, annotated with the objects present on these records, the DL intermediate representations and the generated utterance are available publicly [3]. Since the ontology and the queries are fixed we did not use a reasoner to perform query rewriting. Instead, all the rewritings were computed manually and hard-coded in Google Apps Script to make it available already in the process of time-consuming manual video annotations.

In our experiments, we observed that the size of the raw data records varies a lot between environments, with averages from 12.56 assertions for some environments to 168.85 assertions for others. However, after computing the justifications, on average only 1.68–3.26 assertions remain. Comparing the performance of the natural language generation with several baselines as well as state-of-the-art methods, we observe that all of them benefit from the additional pre-processing step of computing the intermediate representation – with differences up to 37.36

---

[3] `https://cloud.perspicuous-computing.science/s/rPqKAQoWXiq2QSQ`

BLEU points (measuring how close the generated utterances are to the gold standard). A manual evaluation of 100 randomly selected samples also revealed an increased quality of the generated text as well as fewer errors in terms of missing important facts or hallucinated facts. To evaluate the scalability, we also exposed the system to environments not included in the training dataset. While the performance decreases in this setting, again including DL reasoning makes a big difference, as it reduces all data records to a similar form, i.e. it includes only the facts relevant for the handover.

In future work, we would like to make this approach more automated by also learning the TBox (classifying the situations of interest) from the raw data. Additionally, including TBox axioms in the intermediate representation may enable us to generate more detailed and naturalistic explanations for situations that are not time-critical.

References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2 edn. (2007)
2. Chen, Z., Eavani, H., Liu, Y., Wang, W.Y.: Few-shot nlg with pre-trained language model. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 183–190 (2020). https://doi.org/10.18653/v1/2020.acl-main.18
3. Freitag, M., Roy, S.: Unsupervised natural language generation with denoising autoencoders. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018. pp. 3922–3929. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/d18-1426
4. Fuhrman, T., Schneider, D., Altenberg, F., Nguyen, T., Blasen, S., Constantin, S., Waibe, A.: An interactive indoor drone assistant. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 6052–6057 (2019). https://doi.org/10.1109/IROS40897.2019.8967587
5. Lebret, R., Grangier, D., Auli, M.: Neural text generation from structured data with application to the biography domain. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016. pp. 1203–1213 (2016), `http://aclweb.org/anthology/D/D16/D16-1128.pdf`
6. Liu, T., Wang, K., Sha, L., Chang, B., Sui, Z.: Table-to-text generation by structure-aware seq2seq learning. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018. pp. 4881–4888 (2018), `https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16599`

7. Novikova, J., Dusek, O., Rieser, V.: The E2E dataset: New challenges for end-to-end generation. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Saarbrücken, Germany, August 15-17, 2017. pp. 201–206 (2017), `https://aclanthology.info/papers/W17-5525/w17-5525`