Challenges of Military Applications for Data-to-Text Generation

Helen Hastie
School of Mathematical and Computer Sciences
Heriot-Watt University, Edinburgh, UK
h.hastie@hw.ac.uk

Abstract

This position paper describes the potential for data-to-text generation for military applications and the associated challenges in terms of user interaction, system development and deployment. We also present a use-case for a prototype system for debrief generation for autonomous underwater vehicles (AUVs).

1 Introduction

The field of data-to-text generation has the potential to make a huge impact for defence applications, particularly where time-critical, situation awareness is key, be it from missions with one or more autonomous or semi-autonomous vehicles, through to command and control centres where multiple data sources need to be aligned and succinctly summarised. Other applications for data-to-text generation include intelligence gathering through combining, interpreting and summarising disparate sources of intelligence through surveillance or reconnaissance missions. In this paper, we discuss the challenges associated with data-to-text generation for such applications and provide a use-case of an initial system for debrief generation from autonomous underwater vehicles (AUVs) and present some initial results.

2 Challenges: Interacting with the User

Maintaining user trust. For data-to-text technology to be adopted, one needs to ensure that a level of trust is maintained. If this is not the case, users are likely to abandon the system. Trust can be maintained in a number of ways, such as providing further explanations (Nothdurft et al., 2014) or providing back-up information such as graphs along with the generated text (Mahamood et al., 2014) or through multimodal interaction (Daniels et al., 2004). **Information needs**. As has been identified in the literature for medical (Mahamood and Reiter, 2011) and other domains (Gkatzia et al., 2014; Dethlefs et al., 2014), one has to be aware of the different roles of personnel, their varying information needs, communication goals and their individual context. **Managing expectations**. This challenge is particularly prevelant for military spoken dialogue systems (Daniels et al., 2004), where interactive systems are limited in what information they can give. The user must be aware of these limitations in order to minimise frustration. **User-centered design**. Designing data-to-text generation systems requires input from experts either through consultation or sample target text. However, one major challenge is limited access to the military end users. If this is not possible, retired personnel can provide expert advice, however, it is important to get input from a number of personnel for each role, particularly for hand-crafted, expert-driven rule based systems. **Maintaining situation awareness**. During in-mission debriefs, there is a fine balance between maintaining situation awareness of the user through natural language alerts and the risk of distracting the user and adding to cognitive workload. This is a much researched subject in the field of Human-Factors (Endsley, 1995). **Evaluation**. Ideally systems would be evaluated in the situation where they would be ultimately deployed. However, this is likely not to be possible especially for high-stake missions. Evaluation using simulations is a reasonable alternative and assets frequently have some form of simulation available, e.g. for AUVs such as that reported in (Johnson and Lane, 2011). **Evaluation metrics**. Both intrinsic and extrinsic metrics need to be established (Belz and Hastie, 2014). Extrinsic metrics may be trickier to establish as many military applications have long term goals rather than an immediate notion of 'task success' directly related to a generated summary. **Accuracy of generated re-**

ports. Depending on the application, the weighting of importance of precision vs. recall may vary, e.g. for intelligence applications, if there are possible threats one must mention them all (recall), which may outweigh precision, i.e. mentioning items that are irrelevant may have less penalty.

3 Challenges: System Development

Access to data. This is likely the most significant challenge with respect to R&D in this field, given that the majority of missions will be classified. Lack of data is particularly problematic for the development of machine learning techniques. De-classification can take time and in some cases will not be a possibility. One solution, although not ideal, is running missions in simulation, introducing uncertainty, such as adverse environmental conditions and having experts provide debriefs accordingly. **Reliability of data**. One of the main challenges of using real data is the reliability of the sources which may introduce uncertainty into the data, such as break-ups in side-scan sonar imagery or position data with navigation jumps. Data-driven generation methods can be used to capture this uncertainty (Lemon et al., 2010). However, actually knowing that data has a level of uncertainty is important for the end-user, particularly for intelligence applications. How the level of uncertainty of data is reflected in the generated text is an area of further research. **Transferability to new domains/sensors**. In order for this technology to be widely adopted as part of military systems, it is important to develop techniques that allow for the easy transfer of domains and sensors such as from AUVs to surface to airborne vehicles.

4 Challenges: System Deployment

Platforms. Barriers to system deployment that need to be overcome include deploying to systems on various platforms with limited communication bandwidth- this is particularly prevelant for AUVs. **Deployment**. In order for this technology to be transferred to the theatre, one must understand where the technology fits in with standard operating procedures and the acquisition process and have a realistic understanding of where the technology sits on the Technology Readiness Levels (TRLs) (United States Department of Defense, 2011). **Ethics**. There is a risk with any system, that the user may misinterpret the output for any number of reasons. It is, therefore, very important to make data-to-text systems as transparent as possible, leaving an evidence trail as to how text is generated or by providing back-up information in the form of multimodal interaction.

5 Case Study: Data-to-text Generation for AUVs

A prototype system was developed that can automatically generate natural language reports from metadata from AUVs, specifically data exported from the REMUS Vehicle Interface Program (REMUS VIP). The system takes time-series sensor data, mission logs, together with mission plans as its input, and generates post-mission debriefs for human operators in a concise and easy-to-understand manner. The state-of-the-art systems tackling similar issues are designed for simulated environments (Johnson and Lane, 2011) or concentrate on interpreting plans (Tintarev and Kutlak, 2014). To the best of our knowledge, this is the first system that directly deals with noise-prone real data from mission logs combined with mission plans. AUVs are particularly challenging in terms of limited bandwidth for communication so we have concentrated here on post-mission briefs. The architecture includes an Event Identifier module which identifies 3 types of events (lawnmower, traditional and octagonal spiral reacquisitions) combined with various template-based generation modules (e.g. data-error, abort and plan success reporters). We designed the system to be modular for easy addition of new sensors. We tested the Event Identifier on a set of 215 real world REMUS AUV missions conducted by the Heriot-Watt Oceans Systems Lab. These missions were manually annotated with event sequences. Overall event accuracy was 74% with high Precision and Recall for lawnmower scans and traditional reacquisitions. 72% of the missions had all events accurately identified. For more information see (Wang and Hastie, 2015).

6 Discussions and Conclusions

The area of defence provides a rich field of interesting applications and challenges with real potential to create value, improve safety and possibly save lives. In addition, technologies initially deployed for defence purposes are frequently transferred into the mainstream where potential non-military domains include first responders, oil and gas, and agriculture amongst others.

7 Acknowledgements

This research was funded by Dstl under the Autonomous Systems Underpinning Research (ASUR) programme (0913_C2_Ph1_009). I would like to acknowledge Zhuoran Wang for his work on the use-case study project described here. I would like to acknowledge the Oceans Systems Lab at Heriot-Watt University for supplying the mission data and to Prof. Yvan Petillot and Leonard McLean for their consultation and advice. I would also like to acknowledge SeeByte for their advice on this topic.

References

Anja Belz and Helen Hastie. 2014. Towards comparative evaluation and shared tasks for NLG in interactive systems. In Srinivas Bangalore and Amanda Stent, editors, *Natural Language Generation in Interactive Systems*. Cambridge University Press.

Jody J. Daniels, Dan Miksch, and Helen Hastie. 2004. Establishing trust in a deployed spoken language system for military domains. In *Proceedings of AAAI Workshop, the Nineteenth National Conference on Artificial Intelligence, San Jose, CA*.

Nina Dethlefs, Heriberto Cuayahuitl, Helen Hastie, Verena Rieser, and Oliver Lemon. 2014. Cluster-based Prediction of User Ratings for Stylistic Surface Realisation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64.

Dimitra Gkatzia, Helen Hastie, and Oliver Lemon. 2014. Multi-adaptive Natural Language Generation using Principal Component Regression. In *Proceedings of 8th International Natural Language Generation Conference (INLG)*.

Nicholas A. R. Johnson and David M. Lane. 2011. Narrative monologue as a first step towards advanced mission debrief for AUV operator situational awareness. In *Proceedings of ICAR*.

Oliver Lemon, Srini Janarthanam, and Verena Rieser. 2010. Generation under uncertainty. In *Proceedings of the 6th International Natural Language Generation Conference (INLG)*.

Saad Mahamood and Ehud Reiter. 2011. Generating affective natural language for parents of neonatal infants. In *Proceedings of 13th European Workshop on Natural Language Generation (ENLG)*.

Saad Mahamood, William Bradshaw, and Ehud Reiter. 2014. Generating annotated graphs using the NLG pipeline architecture. In *Proceedings of the 8th International Natural Language Generation Conference (INLG)*.

Florian Nothdurft, Felix Richter, and Wolfgang Minker. 2014. Probabilistic human-computer trust handling. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, Philadelphia, PA, U.S.A., June.

Nava Tintarev and Roman Kutlak. 2014. Sassy-making decisions transparent with argumentation and natural language generation. In *Proceedings of the IUI 2014 Workshop on Interacting with Smart Objects*.

United States Department of Defense. 2011. Technology readiness assessment (TRA) guidance.

Zhuoran Wang and Helen Hastie. 2015. A demonstration of a prototype for AUV post-mission debrief generation from metadata. In *Proceedings of AAMAS*.