Data Augmentation with Paraphrase Generation and Entity Extraction for Multimodal Dialogue System

Eda Okur, Saurav Sahay, Lama Nachman

Intel Labs, USA

{eda.okur, saurav.sahay, lama.nachman}@intel.com

INTRODUCTION

- Building the Natural Language Understanding (NLU) module of goal-oriented Spoken Dialogue Systems (SDS) often involves: the definition of intents (and entities), creating domain-specific and task-relevant datasets, annotating the data with intents/entities, iterative training and evaluation of NLU models, and repeating this tedious process for new/updated use-cases.
- This work explores the potential benefits of data augmentation with paraphrase generation for improving the NLU models trained on limited task-specific datasets.
- Our experiments show that paraphrasing using small seed data with model-in-the-loop (MITL) data augmentation strategies helps boost the performance of the Intent Recognition (IR) task.
- We also investigate extracting entities for potentially further data expansion.

MOTIVATION

Fig 1: Learning basic math via play-based multimodal interactions at Kid Space.

- Intelligent conversational agent helping the kids learn 'tens and ones' concepts, along with practicing simple counting, addition, and subtraction operations.
- SDS is a crucial building block for handling efficient task-oriented communication with younger children in play-based learning settings.

MULTIMODAL SDS

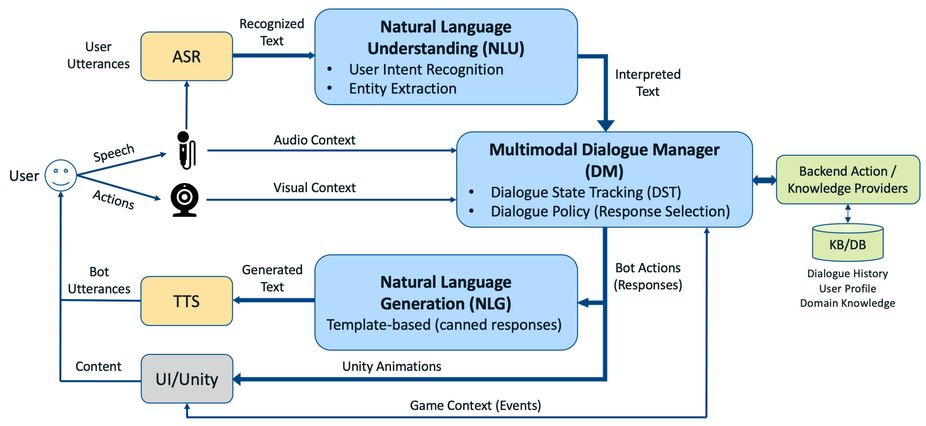

Fig 2: Multimodal SDS pipeline.

METHODOLOGY

Natural Language Understanding (NLU)

- Our NLU and DM models are built on top of the Rasa open-source framework [1].
- The former baseline Intent Classifier in Rasa was inspired by the StarSpace, where embeddings are trained by maximizing the similarity between intents & utterances.
- TF+BERT: In prev. work [2], we enriched this former baseline Rasa NLU architecture by adapting Transformer networks and incorporating pre-trained BERT embeddings.
- DIET+ConveRT: In this work, we adapted the Transformer-based multi-task DIET architecture [3] and incorporated pre-trained ConveRT embeddings [4] to improve Intent Classification performance (and explore the Entity Recognition capabilities).

Paraphrase Generation

- We develop a data augmentation module by training a paraphrasing model to generate paraphrased samples from the original seed utterances to augment the NLU training data.
  - We adapted the BART seq2seq model [5] that we fine-tuned on the back-translated English sentence pairs from the ParaNMT-50M, PAWS, and MSRP corpora.
- We examine the data augmentation with certain heuristics:
  - **baseline (aug3/aug5/aug10)**: augment with 3/5/10x paraphrased samples
  - **inc6low**: paraphrasing only for the low-sample intents or minority classes (<50)
  - **exc5short**: excluding the intents with samples having shorter utterance lengths
- We also investigate model-in-the-loop (MITL) data augmentation techniques using the initial NLU models trained on original samples:
  - **success**: augment only the paraphrased utterances with successful predictions
  - **success_conf90**: success with the confidence level thresholds (>0.9)
  - **all_conf90**: checking the confidence level thresholds only (>0.9)

EXPERIMENTS AND RESULTS

Datasets

- Experiments conducted on Kid Space NLU datasets having utterances from math learning experiences (Planting and Watering activities) designed for 5-to-8 years-old kids [6].
- Some intents are highly generic across usages and activities (e.g., *affirm*, *deny*, *next_step*, *out_of_scope*, *goodbye*), whereas the rest are highly domain-dependent and task-specific (e.g., *intro_meadow*, *answer_flowers/water*, *ask_number*, *counting*).

Statistics/Dataset	Planting	Watering
# distinct intents	14	13
total # samples (utterances)	1927	2115
min # samples per intent	22	25
max # samples per intent	555	601
avg # samples per intent	137.6	162.7
# unique words (vocab)	1314	1267
total # words	10141	10469
min # words per sample	1	1
max # words per sample	74	65
avg # words per sample	5.26	4.95

Table 1: Kid Space NLU Dataset Statistics

Intent Recognition (IR)

Model/Dataset	Planting	Watering
TF+BERT (Baseline)	90.55	92.41
DIET+ConveRT	**95.59**	**97.83**
Performance Gain	+5.04	+5.42

Table 2: NLU/Intent Recognition weighted avg F1 (%): Previous (TF+BERT) and Updated (DIET+ConveRT) Model Results (3 runs of 10-fold CV)

- Adapting the lightweight Dual Intent and Entity Transformer (DIET) architecture with Conversational Representations from Transformers (ConveRT) embeddings significantly improved the IR.
  - Consistent across different use-cases & datasets (Planting and Watering).
- Updated the NLU component in our SDS pipeline by replacing the TF+BERT model with this promising DIET+ConveRT model.

Data Augmentation

Enter text to paraphrase: Oscar has something he wants to say to you children

---

Predictions >>>

Oscar has something to say to you, children.

Oscar has something he wants to tell you, kids.

He has something he wants to tell you, kids.

Oscar has something he wants to tell you, children.

Oscar wants to say something to you, kids.

Fig 3: Example seed and paraphrased utterances.

Method	original	aug3	aug5	aug10
baseline	**95.59**	95.17	94.73	94.75
inc6low	-	95.84	96.06	96.41
exc5short	-	97.70	97.61	97.86
success	-	98.58	98.82	98.65
success_conf90	-	**99.19**	**99.37**	**99.43**
all_conf90	-	98.61	98.75	98.58
Perf. Gain	-	+3.60	+3.78	+3.84

Table 3: NLU/Intent Recognition weighted avg F1 (%) with DIET+ConveRT Models Trained on Augmented Data via Paraphrasing

- **Data Size versus NLU Performance:**

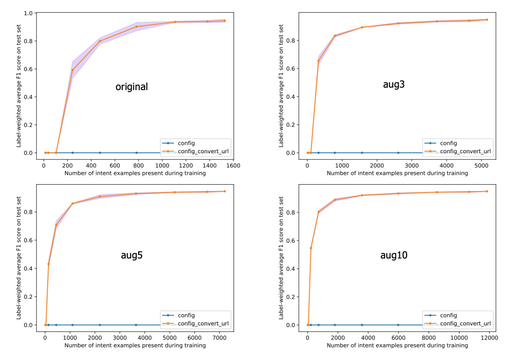

Fig 4: Data size vs. NLU performance for original and augmented datasets.

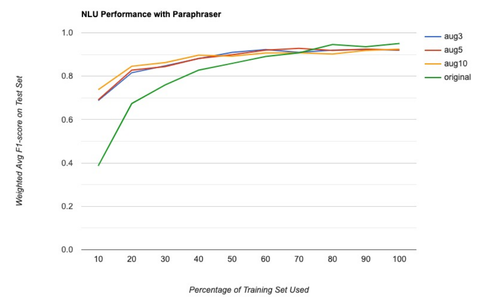

Fig 5: Data percentage vs. NLU performance for original and augmented datasets.

Entity Extraction

	Intent Classification	Entity Recognition
No entities	**95.59**	-
SpaCy NER tagged entities	94.70	72.82
Manually annotated entities	94.91	**97.12**
Auto-annotated entities	94.76	92.64

Table 4: NLU/Joint Intent Classification and Entity Recognition weighted avg F1 (%): DIET+ConveRT Model Results (3 runs of 10-fold CV)

- Entity Recognizers with:
  - SpaCy pre-trained NER tags (generic)
  - Manual annotation for domain-specific entities
  - Auto-annotation for domain-specific entities
- Next: ConceptNet Relatedness Entity Expansion
  - These domain-specific entities would help us achieve lexical entity enrichment via ConceptNet as Knowledge Graph (KG) on original/paraphrased samples.

CONCLUSION

- Focused on improving the NLU module of the task-oriented SDS pipeline with limited datasets.
- Shown that paraphrasing with model-in-the-loop (MITL) strategies using small seed data is a promising approach yielding higher F1-scores for Intent Recognition on task-specific datasets.
- Explored Entity Extraction (for further data expansion and enrichment) to improve the NLU module of our multimodal SDS.
- We believe these results are highly encouraging in our quest to make dialogue systems more robust and generalizable to new intents with limited data resources.

SELECTED REFERENCES

[1] Bocklisch, T., Faulkner, J., Pawlowski, N., and Nichol, A. (2017). *Rasa: Open source language understanding and dialogue management*. CoRR, abs/1712.05181.
[2] Sahay, S., Kumar, S. H., Okur, E., Syed, H., and Nachman, L. (2019). *Modeling intent, dialog policies and response adaptation for goal-oriented interactions*. Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL 2019).
[3] Bunk, T., Varshneya, D., Vlasov, V., and Nichol, A. (2020). *DIET: lightweight language understanding for dialogue systems*. CoRR, abs/2004.09936.
[4] Henderson, M., Casanueva, I., Mrkšić, N., Su, P.-H., Wen, T.-H., and Vulić, I. (2020). *ConveRT: Efficient and accurate conversational representations from transformers*. Findings of the Association for Computational Linguistics: EMNLP 2020.
[5] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2019). *BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2019).
[6] Sahay, S., Okur, E., Hakim, N., and Nachman, L. (2021). *Semi-supervised interactive intent labeling*. Proceedings of the Second Workshop on Data Science with Human in the Loop: Language Advances (DaSH | NAACL 2021). Association for Computational Linguistics.