Reading with Robots: Towards a Human-Robot Book Discussion System for Elderly Adults

Natalie Parde
Human Intelligence and Language Technologies (HiLT) Laboratory
Department of Computer Science and Engineering
University of North Texas

Abstract

As people age, it is critical that they maintain not only their physical health, but also their cognitive health—for instance, by engaging in cognitive exercise. Recent advancements in AI have uncovered novel ways through which to facilitate such exercise. In this thesis, I propose the first human-robot dialogue system designed specifically to promote cognitive exercise in elderly adults, through discussions about interesting metaphors in books. I describe my work to date, including the development of a new, large corpus and an approach for automatically scoring metaphor novelty. Finally, I outline my plans for incorporating this work into the proposed system.

Introduction

Companion robots have received much attention recently as their presence in the commercial marketplace has grown. However, the language technologies guiding their interactions are still nascent. Here, I propose several contributions to natural language processing that will enable a companion robot to engage in more advanced discourse, specifically through discussions of novel metaphors in books. My contributions are driven by the overarching question of whether metaphor can be leveraged as a tool to facilitate cognitive reserve, which gives rise to the following technical questions:

- Can novel metaphoric pairs be reliably extracted from unstructured text (fiction books)?
- What types of features are most beneficial for doing so?
- How can natural-sounding, introspective questions be generated about identified novel metaphors?
- Are companion robots a suitable mode through which to facilitate book discussions with elderly users?

In this work, I describe my work thus far toward answering these questions, and my research plans for the future.

Related Work

Prior work in neuroscience and psycholinguistics has indicated that processing novel metaphors requires additional cognitive effort (Lai, Curran, and Menn 2009), and that older adults experience age-related changes in the way that they process novel but not conventional metaphors (Mashal, Gavrieli, and Kav 2011). Moreover, probable Alzheimer's disease patients suffer no impairments in comprehending conventional metaphors, but struggle to comprehend novel metaphors (Amanzio et al. 2008). These findings make discussions of novel metaphors a particularly intriguing, previously unstudied means of proactive cognitive exercise.

However, although there has been no shortage of *metaphor detection* research to date (that is, a binary decision of whether or not a unit of text is a metaphor), there has been a paucity of research attempting to score *metaphor novelty*, or distinguish novel from conventional metaphors. This has been at least partially due to a lack of novelty-annotated data (Haagsma and Bjerva 2016). Thus, for this thesis I not only establish a benchmark for metaphor novelty scoring, but also publish a large corpus of metaphor novelty-annotated data to stimulate further research in this area.

Finally, various companion robots have been developed for elderly users. A limitation of systems to date has been that none have been able to engage users in dialogue, instead relying on motion (Wada and Shibata 2007) or buttons and canned speech (Tapus, Tapus, and Mataric 2009) for communication. To generate meaningful, cognitively stimulating discussion, more advanced NLP capabilities are needed.

Research Plan

Data Collection

I constructed a new dataset of 18,439 syntactically-related word pairs annotated on a scale from 0 (non-metaphoric) to 3 (highly novel metaphor). The annotations were crowdsourced using Amazon Mechanical Turk, and two trained annotators also labeled each test instance (3162 instances). We (my advisor and I) devised a regression-based approach to aggregate the crowdsourced training data, finding that this technique outperformed traditional aggregation strategies (Parde and Nielsen 2017). We used the adjudicated labels from trained annotators as labels for the test set. Table 1 shows instances from our corpus and their scores.

Metaphor Novelty Scoring

Using our dataset, I trained a deep neural network to predict metaphor novelty scores (Parde and Nielsen 2018). I tested a variety of features for this task, some of which were derived

Instance	Score
Primal Scream: Not a name which would lead you to expect self-absorbed *acoustic doodlings*.	2.5
'Certainly if you *bring* a literal *interpretation* to the story, it's a minefield,' Allan Ahlberg says .	0.5

Table 1: Samples from Our Dataset (Word Pairs Italicized)

Method	*r*	RMSE
Random	0.0048	1.4658
Distribution-Aware Random	0.0007	0.8143
Average	0.0000	0.7192
Ours	**0.4600**	**0.6502**
Tsvetkov et al.	0.2716	0.7804
Ours (AN Pairs Only)	**0.4483**	**0.7312**

Table 2: Metaphor Novelty Scoring Performance

from traditional metaphor detection research, and some of which were newly-designed for metaphor novelty scoring. We compared our approach to several baselines:

- RANDOM: Predicts a random continuous value from 0-3 for each instance.
- DISTRIBUTION-AWARE RANDOM: Learns a probability density function from the training set, and then predicts values based on that distribution for the test instances.
- AVERAGE: Learns the average novelty score from the training set and predicts that value for each test instance.

We also compared our approach to a standard metaphor detection approach, Tsvetkov et al. (2014), modified such that it produced continuous scores rather than binary labels, to experimentally verify that metaphor novelty scoring is a distinct task from regression-based metaphor detection. Tsvetkov et al.'s approach is known for its high performance at metaphor detection (86% accuracy on adjective-noun (AN) pairs); since it was designed for AN pairs, we compare it to a version of our approach also only trained and tested on AN pairs. Our results are shown in Table 2. Our approach outperforms all other approaches, establishing a strong performance benchmark for this task. All results reported are statistically significant ($p < 0.0001$).

Future Work and Proposed Timeline

Several tasks remain as future work. One, currently in progress, is applying the learned scoring model to new books. Another, to be completed in early Spring, involves automatically generating questions about the identified novel metaphors and incorporating them into a dialogue system. I plan to develop a template-based approach based on the Questioning the Author technique, a popular strategy in K-12 education that encourages readers to consider the author's motives in making certain literary decisions (e.g., *"What do you think the author wants us to know by metaphorically saying <novel metaphor >?"*).

The final task will be integrating the finished system with physical robots and conducting a user study to assess the overall performance and likeability of the system. This is projected to be complete by early Summer 2018. Following completion of my degree, I plan to remain in academia and eventually expand upon this project with additional tasks.

Conclusion

In this thesis, I propose the first system specifically designed to promote cognitive exercise in elderly adult users through natural-language interactions with companion robots. To date, I have created a large, publicly available corpus for metaphor novelty, and developed a metaphor novelty scoring system. My planned future work includes applying this system to books, automatically generating introspective questions about the identified novel metaphors, and integrating the system with companion robots to conduct a user study. It is my hope that the contributions of this work will be beneficial not only to the artificial intelligence community, but also to the elderly adults who use the system.

Acknowledgments

This material is based upon work supported by the NSF Graduate Research Fellowship Program under Grant 1144248, and the NSF under Grant 1262860. Any opinions, findings, and conclusions or recommendations are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Amanzio, M.; Geminiani, G.; Leotta, D.; and Cappa, S. 2008. Metaphor comprehension in alzheimer's disease: Novelty matters. *Brain and Language* 107(1):1–10.

Haagsma, H., and Bjerva, J. 2016. Detecting novel metaphor using selectional preference information. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, 10–17.

Lai, V. T.; Curran, T.; and Menn, L. 2009. Comprehending conventional and novel metaphors: An ERP study. *Brain Research* 1284:145–155.

Mashal, N.; Gavrieli, R.; and Kav, G. 2011. Age-related changes in the appreciation of novel metaphoric semantic relations. *Aging, Neuropsych., and Cognition* 18(5):527–543.

Parde, N., and Nielsen, R. D. 2017. Finding patterns in noisy crowds: Regression-based annotation aggregation for crowdsourced data. In *Proceedings of EMNLP*, 1908–1913.

Parde, N., and Nielsen, R. D. 2018. Exploring the terrain of metaphor novelty: A regression-based approach for automatically scoring metaphors. In *Proceedings of AAAI*.

Tapus, A.; Tapus, C.; and Mataric, M. 2009. The role of physical embodiment of a therapist robot for individuals with cognitive impairments. In *Proceedings of RO-MAN*, 103–107.

Tsvetkov, Y.; Boytsov, L.; Gershman, A.; Nyberg, E.; and Dyer, C. 2014. Metaphor detection with cross-lingual model transfer. In *Proceedings of ACL*, 248–258.

Wada, K., and Shibata, T. 2007. Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE Trans. on Rob.* 23(5):972–980.