PuzzLing Machines: A Challenge on Learning From Small Data

Gözde Gül Şahin[†], Yova Kementchedjhieva[α], Phillip Rust[†], Iryna Gurevych[†]
[†]Ubiquitous Knowledge Processing Lab (UKP),
Department of Computer Science, Technical University of Darmstadt
[α]Department of Computer Science, University of Copenhagen
[†]www.ukp.tu-darmstadt.de

Abstract

Deep neural models have repeatedly proved excellent at memorizing surface patterns from large datasets for various ML and NLP benchmarks. They struggle to achieve human-like thinking, however, because they lack the skill of iterative reasoning upon knowledge. To expose this problem in a new light, we introduce a challenge on learning from small data, *PuzzLing Machines*, which consists of *Rosetta Stone* puzzles from Linguistic Olympiads for high school students. These puzzles are carefully designed to contain only the *minimal* amount of parallel text necessary to deduce the form of unseen expressions. Solving them does not require external information (e.g., knowledge bases, visual signals) or linguistic expertise, but meta-linguistic awareness and deductive skills. Our challenge contains around 100 puzzles covering a wide range of linguistic phenomena from 81 languages. We show that both simple statistical algorithms and state-of-the-art deep neural models perform inadequately on this challenge, as expected. We hope that this benchmark, available at https://ukplab.github.io/PuzzLing-Machines/, inspires further efforts towards a new paradigm in NLP—one that is grounded in human-like reasoning and understanding.

1 Introduction

Kahneman (2011) discusses the two modes of human thinking which perfectly encapsulate the current (so called System1) and the desired state (System1+System2) of the deep learning field. System1 handles tasks that humans consider fast, intuitive and automatic, such as object detection and document classification. Recent deep learning (DL) models have shown great promise at this type of tasks—thanks to large training datasets. Yet, it is through slow, rational and sequential mechanisms that human-like abstract reasoning happens, to enable learning from just a few examples. This System2-style modeling is still in its early stages in DL, but is recognized as a much needed next step in the field (McClelland et al., 2019; Marcus, 2020; LeCun, 2020; Bengio, 2020). To foster research in this promising direction, we propose a unique challenge on "learning from small data": *PuzzLing Machines*, based on the Linguistic Olympiads—one of the 13 recognized International Science Olympiads targeted at high-school students.

Chikasaw	**English**
1. Ofi'at kowi'ã lhiyohli.	The dog chases the cat.
2. Kowi'at ofi'ã lhiyohli.	The cat chases the dog.
3. Ofi'at shoha.	The dog stinks.
4. Ihooat hattakã hollo.	The woman loves the man.
5. Lhiyohlili.	I chase her/him.
6. Salhiyohli.	She/he chases me.
7. Hilha.	She/he dances.
Now you can translate the following into Chickasaw:	
	The man loves the woman.
	The cat stinks.
	I love her/him.
Translate the following into English:	
Ihooat sahollo.	
Ofi'at hilha.	
Kowi'ã lhiyohlili.	

Table 1: The "Chickasaw" puzzle (Payne, 2005)

The *PuzzLing Machines* challenge is based on one of the most common puzzle types in the Linguistic Olympiads: the *Rosetta Stone* puzzles (Bozhanov and Derzhanski, 2013), a.k.a. translation puzzles. An example is given in Table 1.[1] Although these puzzles take the form of a traditional "machine translation" task, they are different in many ways: Rosetta Stone puzzles contain a minimal, carefully designed set of parallel expressions (words, phrases or sentences) in a for-

[1]Copyright University of Oregon, Department of Linguistics.

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1241–1254