Shane Arwood
November 13, 2022
Dr. Karen Mazidi
CS 4395.001

# CS 4395: ACL Paper Summary

## a. Title, author list, and affiliations

The paper is titled “TellMeWhy: A Dataset for Answering Why-Questions in Narratives” [1]. The authors are as follows:

- Yash Kumar Lal of Stony Brook University
- Nathanael Chambers of the US Naval Academy
- Raymond Mooney of the University of Texas, Austin
- Niranjan Balasubramanian of Stony Brook University

Yash Kumar Lal, the lead author, is a PhD student in Computer Science that is focused on natural language understanding in stories. He is working with two professors in this exploration: Nathanael Chambers and Raymond Mooney. Nathanael Chambers researches NLP with a focus on understanding the world from texts and serves as Co-Director of the US Naval Academy’s Center for High Performance Computing [2]. Raymond Mooney is the Director of the UT Artificial Intelligence Laboratory and is currently focusing his research in NLP and computational linguistics [3]. Yash Kumar Lal is being advised by Niranjan Balasubramanian, Assistant Professor of the Department of Computer Science at Stony Brook University whose research interests lie in the challenge of systems that can reason in natural language texts [4].

## b. Summary of problem addressed by the paper

The paper aims to measure how well several models perform in answering why-questions in narratives. It introduces a dataset called TellMeWhy, which contains 30k questions and answers to why questions in short narratives [1]. The paper discusses how why-questions were formulated and how the TellMeWhy dataset was created using human evaluation, explains why why-questions are difficult for NLP models to answer, and describes several existing question-answering models. The efficacy of these models, specifically GPT-2 and UnifiedQA, were determined using the TellMeWhy dataset and actual human-evaluation. The goal of the paper was to bring attention to the answering of why-questions in the field of natural language processing.

The authors concluded that state-of-the-art question-answering models do not perform nearly as well as human evaluators on why-questions. This was especially true

for questions that required answers that could not be found explicitly in the text and relied on common external knowledge for the answer to be inferred. They call for more research to be done on incorporating common-sense knowledge and reasoning capabilities into question-answering models, as this seems to be the best route for addressing the scarcity of progress in getting models to answer why-questions.

### c. Summary of prior work

The lead author is interested in reasoning in narratives, or stories. This often requires analyzing why characters in a narrative do what they do, but such why-questions require the piecing together of common knowledge as well as disjoint parts of a narrative. This makes such questions extremely difficult for NLP models to address; there is not an explicit answer to be found in the text.

Prior work in this domain is limited. The few datasets that contain why-questions are all either extremely small, not focused on narratives, or focused on connecting a series of events rather than reasoning why certain events have happened [1]. In 2008, an existing question-answering system was built upon to include the answering of why-questions using a corpus, called NAZEQA [5]. Unfortunately, the dataset was extremely small. NarrativeQA, a dataset that does include the use of stories, is only around 10% why-questions [6].

Even when work is done with why-questions, they are not necessarily the type of why-questions of interest; they focus on extracting information rather than reasoning information [7]. It is then evident that while there is some prior work in models answering why-questions and models answering questions from narratives, the intersection between the two has not been dealt with extensively yet.

### d. Unique contributions of this paper

This paper is unique in that it has not only developed an extensive dataset of why-questions and answers, but tested them against state-of-the-art models to demonstrate a specific area of improvement in NLP research that is often overlooked. It focuses on reasoning in narratives specifically. This is important because reasoning is a critical part of natural language understanding, and being able to apply it in narratives would expand the capabilities of natural language processing to another critical type of text: stories. However, this reasoning requires not only knowledge extraction from text, but the integration of common knowledge as well. This would be an extremely useful skill for natural language processing models to implement, but not a lot of prior work has been done on the topic. By demonstrating the first step in developing a strong model, creating the dataset to test why-questions, the paper opens the door for more research to be done in this area.

## e. How the authors evaluated their work

The authors' work is a segment within natural language processing that does not have many automation tools readily available. To evaluate their work and ensure the creation of the dataset, as well as its analysis for QA models, was high-quality, the authors had to use crowdsourcing. They did this via Amazon Mechanical Turk [1]. The authors put a great amount of thought into the details going into crowdsourcing, such as giving annotators the opportunity to rank different questions and answers on validity and grammaticality [1], which came out to be 95% and 99% respectively [1]. They were also asked to provide whether answers were found in the text implicitly or explicitly, and several rounds of crowdsourcing was done to achieve this. The authors also used human evaluators for their testing of the QA models against the dataset. This was required due to this work being some of the first of its kind; humans are still the only ones who reliably answer why-questions to allow the authors to evaluate their work.

## f. Conclusion of author citations and value of their work

The number of citations for each author is as follows (according to Google Scholar):

- Yash Kumar Lal - 149
- Nathanael Chambers - 6578
- Raymond Mooney - 43147
- Niranjan Balasubramanian - 2853

Each of the authors has done important work in Natural Language Processing and Machine Learning, especially in extracting knowledge from texts like stories. Yash Kumar Lal in particular, while still a student with few citations, is exploring an extremely interesting, multi-faceted section of NLP: reasoning in stories. It is attempting to address a very human skill: piecing together knowledge from life experiences and connecting them to stories to understand why events in the story might have happened. It seems extremely difficult to implement in NLP, and this paper is valuable for describing it so well and taking steps to contribute to its future.

Raymond Mooney, the author with the most citations by far, is working on this specific problem with Yash Kumar Lal. Raymond Mooney's career has been expansive across many societies and affiliations, and he has addressed many research areas [3], but his work in reasoning stands out. He has made significant contributions to Explainable AI, Semi-supervised Learning, Transfer Learning, and Abductive Reasoning. Humans take reasoning for granted despite it being vital to our learning and processing of language. That he has identified the need for more growth in this field and contributed so much research to it is commendable.

## References


[1] Yash Kumar Lal, Nathanael Chambers, Raymond Mooney, and Niranjan Balasubramanian. 2021. TellMeWhy: A Dataset for Answering Why-Questions in Narratives. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 596–610, Online. Association for Computational Linguistics.

[2] “Nate Chambers.” *Nate Chambers : US Naval Academy*, https://www.usna.edu/Users/cs/nchamber/.

[3] “Raymond J. Mooney.” *Raymond J. Mooney's Home Page*, https://www.cs.utexas.edu/~mooney/.

[4] Balasubramanian, Niranjan. *Niranjan's Home Page.*, https://www3.cs.stonybrook.edu/~niranjan/research.html.

[5] Ryuichiro Higashinaka and Hideki Isozaki. 2008. Corpus-based Question Answering for why-Questions. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.

[6] Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, Edward Grefenstette; The NarrativeQA Reading Comprehension Challenge. *Transactions of the Association for Computational Linguistics* 2018; 6 317–328. doi: https://doi.org/10.1162/tacl_a_00023

[7] Joanna Mrozinski, Edward Whittaker, and Sadaoki Furui. 2008. Collecting a Why-Question Corpus for Development and Evaluation of an Automatic QA-System. In *Proceedings of ACL-08: HLT*, pages 443–451, Columbus, Ohio. Association for Computational Linguistics.