

Literature review

Title: Solving math word problems with controlled natural language

Author: Fabian Afatsawo - 12105155

Supervisor: Giovanni Sileno

Keywords: Controlled Language, Controlled Natural Language, Natural Language Processing, Math word problems, Formal Logic, First-order Logic

Problem: Since the invention of the calculator, computers do an excellent job in calculating math problems. The language used in math is a substrate of natural language (NL), using technical terms and grammatical conventions. Once given to a computer, it is converted to logic. One could say, the language of math is a subset of NL with a set of rules. In that sense, it is a controlled natural language (CNL). A CNL is a subset of NL that is human and machine-understandable. But, when math is expanded to NL computers have a hard time understanding it. To this day, simple middle school word math problems are difficult to solve for a computer. Attempts have been made by using autoregressive language models like GPT-3 or converting NL straight into logic. Unfortunately, the results are still unsatisfiable. To tackle this, conversion of NL to CL might solve the problem. Therefore, can NL be converted into a CL that in turn is able to solve simple math word problems by using logic? If so, from simple math the understanding can expand to more advanced math. One can imagine a computer being able to verify proofs. Not only that but formalizing proofs itself.

~~Draft: Can simple math word problems be solved by utilizing a controlled natural language? Can natural language be converted into a controlled natural language? In addition, can a controlled natural language be utilized to solve simple math word problems?~~

Literature:

[1] : In this paper, an attempt is being made to rewrite a human-oriented CNL into a machine-oriented CNL. The paper suggests that this transformation could unlock This paper explores the feasibility of rewriting a human-orientated CNL represented in Simplified English into a well known machine-oriented CNL represented in ACE CNL. According to the paper, this could unlock significant silos of general-purpose domain knowledge. By using NLP techniques, the text was analyzed and ACE compliant sentences were generated. As a validation check, the ACE APE parser was used. Finally, the ACE compliant sentences were used for knowledge extraction by extracting DRS and OWL triples. The initial results showed that simplified text can be rewritten into a CNL and can be used for knowledge extraction. This is relevant to the problem because it shows a possible solution for solving the math word problems (NL) to CL part.

[2]: In this paper, an attempt is being made to move natural language legal text to their respective set of machine-readable conditions. The paper proposes the use of NLP techniques towards extracting rules of legal texts. According to the paper, a combination of extracting syntax and linguistic information leads to a powerful solution. The initial results are

still unsatisfiable. With respect to the problem, this can be a possible approach to convert math word problems (NL) to CL.

[3]: In this paper, an argument is made on developing a controlled natural language for mathematics. The paper points out that there is little consensus on the issue, and mathematicians and computer scientists tend to line up on opposite sides. Technology is still far from making a semantic reading of mathematics as it is currently written. Machine learning techniques, as well as linguistic approaches, have not yet reached the ability. Mathematicians are still far from the mass adoption of proof assistants. Adoption has been gradual, and structural reasons hinder the adoption of proof assistants. The paper points out that there is value in bridging the above. CNL technology can help to bridge the gap. This is relevant because it is the exact problem that our paper is trying to solve.

[4] In this paper, the capability of neural models on the task of parsing First-Order Logic (FOL) from natural language sentences is examined. Analysis showed difficulties faced by neural networks in modelling FOL and ways to tackle them. This work goes from NL to Logic. This is relevant to the problem because we start with word problems (NL), convert to CL, extract Logic and then make a calculation on this logic.

[5] In this paper, a first step is taken towards automatic formalisation and verification of textbook mathematical text. Therefore, a controlled natural language for mathematics (CLM) is developed. This CLM is a subset of English with restricted grammar and a dictionary. Second, the CLM automatically gets translated into a system-independent formal language (MathAbs). The paper claims that there is no system that provides such linguistic features as their system does. However, the coverage of CLM facilitates common reasoning patterns but is limited. Furthermore, theorem provers need improvement. This is relevant to the problem because it tries to solve the same problem in a sense. Both approaches try to solve textbook math problems using CL.

Bibliography:

- [1] Safwat, H., Zarrouk, M., & Davis, B. (2018). Rewriting simplified text into a controlled natural language.
- [2] Dragoni, M., Villata, S., Rizzi, W., & Governatori, G. (2016). Combining NLP Approaches for Rule Extraction from Legal Documents. In *Proceedings of the workshop on "Mining and Reasoning with Legal Texts" collocated at the 29th International Conference on Legal Knowledge and Information Systems*.
- [3] Hales, T. (2019). An Argument for Controlled Natural Languages in Mathematics.
- [4] Singh, H., Aggrawal, M., & Krishnamurthy, B. (2020). Exploring neural models for parsing natural language into First-Order Logic. *arXiv preprint arXiv:2002.06544*.
- [5] Humayoun, M., & Rafalli, C. (2008). MathNat-Mathematical Text in a Controlled Natural Language. Special issue: Natural Language Processing and its Applications. *Journal on Research in Computing Science*, 46.

All possible references¹:

Name	Description	Source	Author
Controlled English for knowledge representation	The definition of Controlled Natural Languages (CNL). Furthermore, exploring the use of CNL to make knowledge representations. ¹	https://www.zora.uzh.ch/id/eprint/33191/10/documental_thesis_kuhn.pdf	Kuhn, T
Controlling Controlled English. An Analysis of Several Controlled Language Rule sets.	Analysis on eight different Controlled English rule sets.	https://aclanthology.org/2003.eamt-1.12.pdf	Sharon O'Brien
Attempto Controlled English (ACE)	Creation of Attempto Controlled English (ACE)	https://arxiv.org/pdf/cmp-ig/9603003.pdf	Norbert E. Fuchs, Rolf Schwitter
Attempto controlled English for knowledge representation	Attempto controlled English (ACE) is a controlled natural language, a subset of English that is human and machine-understandable. The Attempto Parsing Engine (APE) translates ACE into Discourse Representation Structures (DRS, first-order logic).	https://www.zora.uzh.ch/id/eprint/8110/2/veniceV.pdf	Fuchs, N E; Kaljurand, K; Kuhn, T
The Controlled Natural Language of Randall Munroe's Thing Explainer	Analysis of Randall Munroe's book 'Thing Explainer' written in a controlled natural language.	https://arxiv.org/pdf/1605.02457.pdf	Kuhn, T
Rewriting Simplified Text into a Controlled Natural Language	An attempt to convert simple English text into ACE. This is done by extracting the syntax (POS tags) and then applying ACE rules. Thereafter, alternative sentences will be created on which APE will be checked.	doi:10.3233/978-1-61499-904-1-85 CNL-2018	Hazem Safwat, Manel Zarrouk, Brian Davis
Combining NLP Approaches for Rule Extraction from Legal Documents	Attempt to convert natural legal text into the respective set of machine-readable conditions. This is done by combining different NLP techniques towards the extraction of rules from legal documents.	doi:10.1007/978-3-030-00178-0_19	Mauro Dragoni; Serena Villata; Williams Rizzi; Guido Governatori
Training Verifiers to Solve Math Word Problems	An attempt to solve of 8.5K high quality linguistically diverse grade school math word problems. With the use of GPT-3	https://arxiv.org/pdf/2110.14168.pdf https://openai.c	Karl Cobbe; Vincent Kosaraju; Mohammad Bavarian;

¹ Bold names are used in the bibliography

		om/blog/grade-school-math/?s=09	Jacob Hilton; Reiichiro Nakano; Christopher Hesse; John Schulman
An Argument for Controlled Natural Languages in Mathematics	Explaining the need for Controlled Natural Language in Mathematics. With objections by mathematicians.	https://jiggerwit.wordpress.com/2019/06/20/an-argument-for-controlled-natural-languages-in-mathematics/	Thomas Hales
MathNat - Mathematical Text in a Controlled Natural Language	The MathNat project aims at being a first step towards automatic formalisation and verification of textbook mathematical text.	https://www.researchgate.net/publication/228571904_MathNat_Mathematical_Text_in_a_Controlled_Natural_Language	Muhammad Humayoun and Christophe Raffalli
Exploring Neural Models for Parsing Natural Language into First-Order Logic	An attempt to parse Natural Language into First-Order Logic using Neural Models.	https://arxiv.org/pdf/2002.06544v1.pdf	Hrituraj Singh; Milan Aggrawal; Balaji Krishnamurthy
Extracting Formal Specifications from Natural Language Regulatory Documents	Formal verification techniques provide a way to determine whether regulatory documents are consistent and whether implementations conform to them. To apply these techniques a formal description of the regulation needs to be extracted. We present a framework, under which NLP techniques can be brought to bear, to aid a requirements engineer in extracting the formal description.	https://aclanthology.org/W06-3902.pdf	Nikhil Dinesh; Aravind Joshi; Insup Lee
Natural- To Formal-Language generation using Tensor Product Representations.	An attempt to parse natural language into formal language using tensor product representations.	https://openreview.net/pdf?id=ByIPSkHKvB	Anonymous authors