Extracting and evaluating educational concept dependencies from Large Language Models

Dominik Glandorf
Matrikelnummer 6007407
dominik.glandorf@student.uni-tuebingen.de

Anastasiia Alekseeva
Matrikelnummer 5994775
anastasiia.alekseeva@student.uni-tuebingen.de

GitHub repository: https://github.com/mlcolab/learning-dependencies

Abstract

1 Introduction

Large Language Models (LLMs) are trained on immense corpora of text and have proven to build on embedded factual information when performing well in downstream tasks such as question answering. Accessing this knowledge, which is represented by billions of parameters and the network's architecture, has given rise to the research field of Knowledge Extraction from LLMs (Cohen et al.), which rests on the assumption that the language models can help retrieve information on the relation between entities.

This work tries to contribute to the research field and to answer the question of whether general LLMs are able to provide a valid dependency structure of educational concepts. To find the answer to the research question, we are going to solve two main subtasks: 1) extraction of implicit educational concept dependencies, and 2) evaluation of the retrieved dependencies or the resulting graphs by existing knowledge media.

Apart from the field of knowledge extraction from LLMs, which is a subfield of computational linguistics, our work may contribute to applied education sciences by improving computer-assisted education and creating new perspectives for a subfield of knowledge engineering that has rather stagnated in the previous years.

In particular, effective and efficient instruction does not only incorporate what to teach but also how to teach, especially the order of instruction. It is proven that *instructional sequencing* enhances the process of learning when prerequisites of educational content are known to the student or taught first [Morrison et al., 2019]. Within educational content, Merill (1983) differentiated facts, concepts, principles, rules, procedures, interpersonal skills, attitudes, and their sole recall from their application. We will focus on concepts and their relations defined understanding.

To be more precise, if one concept is a prerequisite of another, we refer to this relation as a concept dependency. For example, to understand the concept of a derivative, having knowledge about the concept of a function will facilitate or even enable learning. When the dependencies are thought of as directed edges between nodes that represent concepts, a concept dependency graph emerges which is a special type of a knowledge graph [Wang et al., 2016]. This graph is also called *concept map* in the field of Learning Sciences. The graph can be used to advance curriculum planning (Yang et al. [2015]), especially for new topics that might not be covered in textbooks or automated assessment (Wang, 2015).

To conclude, our main contributions are the following. First, there is little research on how to evaluate the precision of extracted concept dependencies. Therefore, we propose a set of methods to use existing unstructured knowledge sources, namely Wikipedia and textbooks, to create baselines for evaluation. Due to the heuristic character of these methods, we conducted a manual assessment to test their suitability for our purpose. The resulting dataset can be used as a baseline for further research.

Second, the emerging field of prompt engineering provides a constellation of commands to query LLMs, the majority of which are experimental and cannot be considered fully reliable. We propose a method called *output refeeding* that allows mining the educational concept dependencies by sequentially querying the language generation model and transforming its answers into a knowledge graph.

1.1 Related work

Talukdar and Cohen [2012] defined the prerequisite relation in terms of the consumption of information about concepts. Vuong (2011) if a better graduation rate given prerequisite knowledge is fulfilled. Concepts are often equated with Wikipedia articles (Talukdar and Cohen, 2012; Wang, 2015).

Prerequisites can be inferred from learner behavior by testing their performance after being presented different instructional sequences (Pavlik et al., 2008, Vuong et al., 2011). However, this has the disadvantage of disengaging users with too difficult concepts before teaching easier or necessary ones. Experts usually dispose of the required knowledge about concepts to create concept maps. The high cost of expert knowledge motivates the automated extraction of concept dependencies from appropriate sources.

Multiple approaches of prompt engineering for graph creation have been recently developed. Few-shot prompting proved successful (Cohen et al.)

2 Method

In this section, we will first detail our research design and then the characteristics of our information sources as well as the methods that we used to produce the knowledge graph.

2.1 Research design

Our research lies at the intersection of education sciences and computational linguistics that define the choice of methodology. On the one hand, we prompt a state-of-the-art large language model to build entity dependency graphs and evaluate its performance on the concept relations retrieved from Wikipedia and textbooks. On the other hand, we should ensure that the output is sound in terms of educational value, which requires additional qualitative assessment.

Since no ground-truth knowledge graph exists and cannot be achieved in the field of education, we choose textbooks and Wikipedia as the baseline to be able to estimate the precision of the dependencies mined from the LLM. Each information extraction method is not fully independent of the other, in particular, entity recognition and concept disambiguation steps in the procedure for LLM and textbooks depend on Wikipedia. The evaluation was performed via 1) inspecting the output in the interactive dashboard, 2) manual labeling, and 3) convergence between the LLM and two baselines metric.

2.2 Baseline extraction

2.2.1 Wikipedia

Wikipedia is a valuable resource that has a good structure of concepts. One of the key factors that contribute to it is the use of a hierarchical structure. Articles are categorized into broader topics, and those topics are further divided into subtopics. This structure helps readers to navigate through the information in a logical and organized manner, allowing them to quickly find the information they need. Another factor is the use of hyperlinks to connect related articles. This allows readers to move seamlessly between articles and explore related concepts without having to leave the site or perform additional searches. Wikipedia also has a strong editorial process that ensures that articles are well-researched, well-written, and up-to-date. This process helps to maintain the quality and

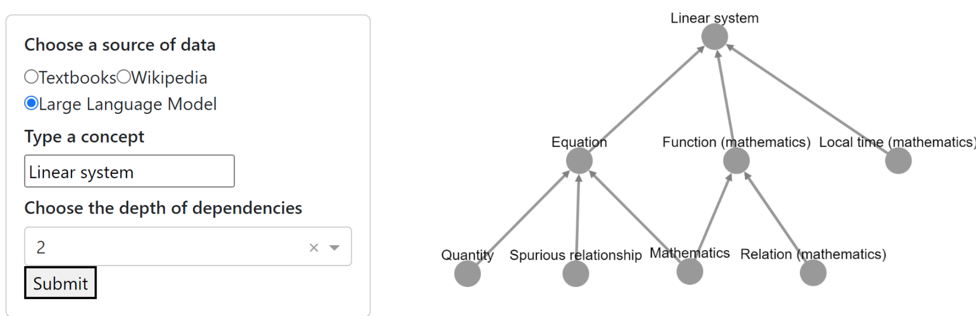

Figure 1: A screenshot of the intercative application.

accuracy of the information presented on the site, which in turn helps to reinforce the logical structure of the concepts presented.

2.2.2 Textbooks

We chose textbooks on linear algebra as they tend to have a good structure and order compared to other textbooks because linear algebra is a well-established and mature field of mathematics with a clear and logical progression of concepts. Moreover, linear algebra is used extensively in other fields, such as engineering, physics, and computer science. Therefore, many textbooks on linear algebra are written with the needs of these fields in mind, which helps to structure the content of the textbook in a clear and practical manner. Finally, the concepts in linear algebra are interrelated, and a strong understanding of earlier topics is often necessary to understand later topics.

We used 10 books that are free and available online.

First, the textbooks were converted from pdf format to raw texts and went through basic preprocessing. Then, Wikifier [Brank et al., 2017] was exploited for entity recognition and disambiguation.

2.3 LLM extraction

2.4 Manual inspection via Dashboard

We developed a Dashboard via graphing library for Python Plotly[1] (Fig.1). The application is devised so that one could examine the resulting graph itself in a convenient environment. It allows plotting the concept dependencies graphs for three knowledge sources or types of dependency mining methods (LLM, Wikipedia, and textbooks) for different levels of dependencies.

[1]https://plotly.com/python/

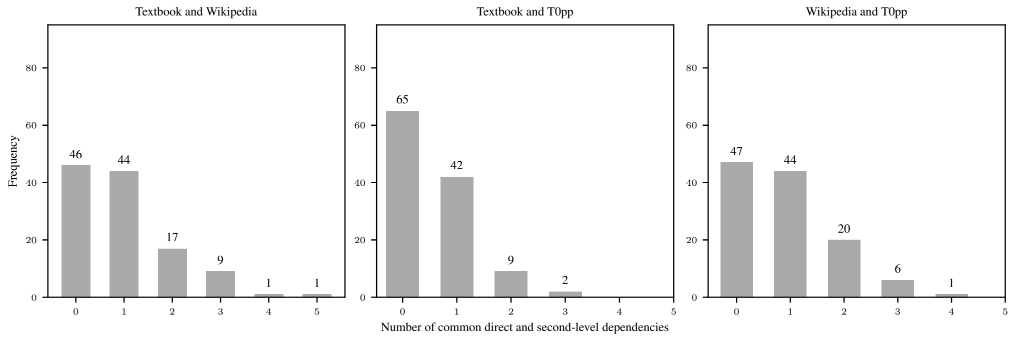

Figure 4: Convergence metrics. Number of common direct dependencies between sources.

2.5 Manual baseline

2.6 Convergence statistics

3 Results

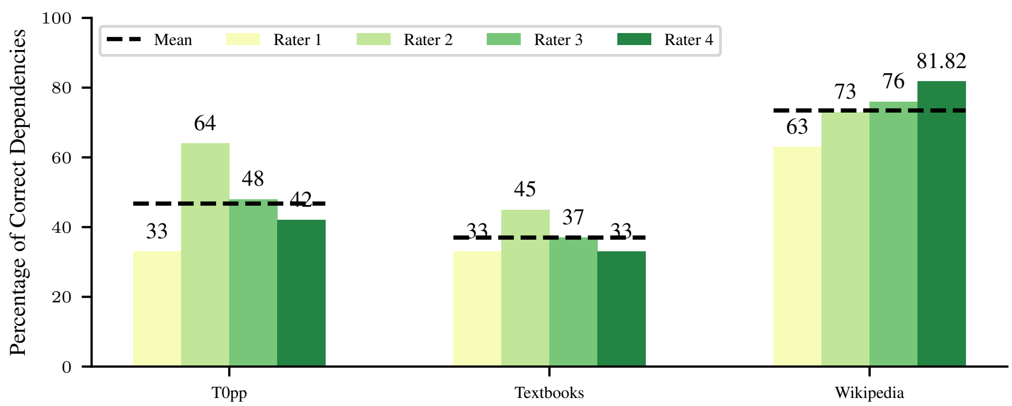

Figure 2: Rating.

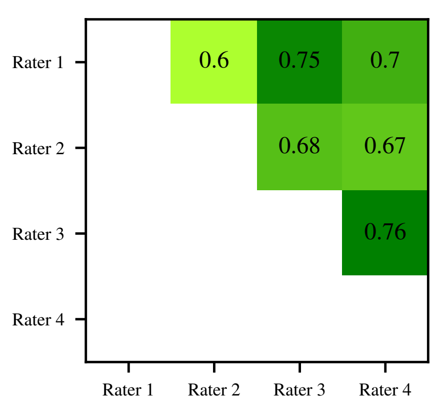

Figure 3: Interrater reliability, kappa coefficient.

4 Discussion

As shown, current LLMs are not sufficiently precise in creating concept maps, which confirms previous research Hwang et al. [2021]. This could be improved by training the pre-trained models on the knowledge graph data mined from Wikipedia or any other reliable sources of concept dependencies in a specific subject of instruction West et al..

References

Roi Cohen, Mor Geva, Jonathan Berant, and Amir Globerson. Crawling the internal knowledge-base of language models. URL `https://arxiv.org/abs/2301.12810`.

Gary R Morrison, Steven J Ross, Jennifer R Morrison, and Howard K Kalman. *Designing effective instruction*. John Wiley & Sons, 2019.

Shuting Wang, Alexander Ororbia II, Zhaohui Wu, Kyle Williams, Chen Liang, Bart Pursel, and C Lee Giles. Using prerequisites to extract concept maps from textbooks. In *Proceedings of the 25th acm international on conference on information and knowledge management*, pages 317–326, 2016.

Y. Yang, H. Liu, J. Carbonell, and W. Ma. Concept graph learning from educational data. *in WSDM*, pages 159–168, 2015.

Partha Talukdar and William Cohen. Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 307–315, 2012.

Janez Brank, Gregor Leban, and Marko Grobelnik. Annotating documents with relevant Wikipedia concepts. *Proceedings of SiKDD*, 472, 2017.

Jena D. Hwang, Chandra Bhagavatula, and et al. (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 6384–6392, 2021.

Peter West, Chandra Bhagavatula, and et al. Symbolic knowledge distillation: from general language models to commonsense models. URL `https://arxiv.org/abs/2110.07178`.

5 Appendix

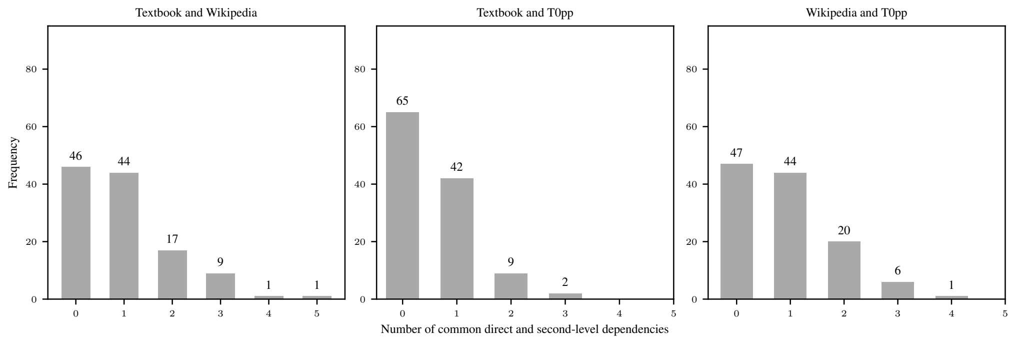

Figure 5: Convergence metrics. Number of common direct and second-level dependencies between sources.