

(続紙 1)

京都大学	博士 (情報学)	氏名	劉 倩瑩 (Liu Qianying)
論文題目	Numerical Reasoning in NLP: Challenges, Innovations, and Strategies for Handling Mathematical Equivalency (自然言語処理における数値推論: 数学的同等性の課題、革新、および対処戦略)		
(論文内容の要旨)			
<p>Numerical reasoning holds a pivotal role in both the realm of natural language processing (NLP) research and practical applications. In the context of NLP, numerical reasoning is a crucial component, playing a pivotal role in pursuing Artificial General Intelligence (AGI). It serves as a vital metric for evaluating models' reasoning abilities, enabling them to comprehend, generate, and analyze data across diverse domains such as science, finance, and news. This application of numerical reasoning drives progress in language translation, sentiment analysis, and text summarization. Beyond research, numerical literacy is indispensable in areas like personal finance, data analysis, and problem-solving, equipping individuals to navigate modern complexities. This thesis delves into the significance of numerical reasoning and its trajectory in research, introducing pertinent tasks and datasets. The central focus lies in tasks with mathematical equivalency as a challenge. Key challenges are identified and addressed, including the costly expert annotation dilemma, mismatch of training objective in supervised training, and enormous search space and false-matching in weak supervision.</p> <p>Chapter 1 gives an in-depth overview of numerical reasoning, emphasizing its importance across domains. Historical context illuminates progress made and remaining gaps.</p> <p>Chapter 2 examines the two types of numerical tasks: Math Word Problem (MWP) and Numerical Question Answering (NQA), establishing a foundation to grasp inherent complexities. Building on prior research, the thesis spotlights a core feature—multiple equivalent expressions in answer equations, denoted as mathematical equivalency. This presents challenges for effective problem-solving.</p> <p>Chapter 3 presents a novel data augmentation method that reverses the mathematical logic of MWPs to produce new high-quality math problems and introduce new knowledge points that can benefit learning the mathematical reasoning logic. Experimental results on the Chinese dataset Math23K and English dataset AllArith show that our method can gain improvement over the back translation augmentation baseline.</p> <p>Chapter 4 presents three pretraining tasks that operate at both the whole program and sub-program level for solving financial analysis NQA tasks. This</p>			

guides the model to focus on useful variables and encourages the model to identify key evidence. Experiments on two financial analysis datasets (FinQA and MultiHiertt) and two pretrained language baselines (BERT-Base and RoBERTa-Large) all gain more than 3.3% accuracy improvement against the baseline.

Chapter 5 introduces a textually enhanced contrastive learning framework for solving MWPs which forces the models to distinguish semantically similar examples while holding different mathematical logic. A self-supervised manner strategy is adopted to enrich examples with subtle textual variance by textual reordering or problem reconstruction. The hardest example is retrieved to differentiate samples from both equation and textual perspectives and guide the model to learn their representations. The method achieved state-of-the-art performance on benchmark Chinese dataset Math23K and English dataset Asdiv-A, and gains accuracy improvement of more than 10% on adversarial challenge test sets that test the robustness of the model.

In Chapter 6, a controlled equation generation solver for solving MWPs is proposed by leveraging a set of control codes to guide the model to consider certain reasoning logic and decode the corresponding equations transformed from the human reference. Experimental results show the effectiveness of the method by outperforming strong baselines on the single-unknown problem (Math23K) and the multiple-unknown problem (DRAW1K, HMWP) datasets.

Chapter 7 presents a novel search algorithm with the combinatorial strategy for solving MWPs in a challenging weakly supervised setting with only answer values given (ground-truth equations are not allowed to use). The method can compress the search space by merging mathematically equivalent equations. The compression allows the search algorithm to enumerate all possible equations and obtain high-quality data. The noise in the pseudo labels that hold wrong mathematical logic is investigated and a ranking model to denoise the pseudo labels is proposed. The ablation study reveals that the high-quality pseudo data can be efficiently searched for training by our method. Experimental results on Math23K datasets show that the method outperforms two strong baselines with more than 4% answer accuracy and reduces the false-matching noise.

Chapter 8 concludes the Mathematical Equivalency centric approaches introduced in this thesis and discusses the directions of future research.

(続紙 2)

(論文審査の結果の要旨)

本論文は、自然言語で記述された数値推論に関する文章題に関して、解答形式における複数の等価表現（数学的等価性）の識別と分析に焦点を当て、数値推論の高度化を行ったものである。本研究の成果は以下の通りである。

1. 数値推論における数学的等価性の問題から、専門家のアノテーションには大きなコストがかかるという課題があり、学習データに対する効率的でスケーラブルなアノテーション手法の開発が必要となる。本論文では、数式の反転を利用して、新たな高品質な数学問題を生成する手法を提案した。さらに、数式全体と数式の一部に関して設計された3つの事前学習タスクを導入し、適切な変数に焦点をあてて重要な特徴を学習するようにモデルを導く手法を考案し、標準的な数学推論タスクにおいて精度改善を実現した。
2. 教師あり学習におけるノイズの問題に対処するため、モデルの頑健性を高める方法が研究されている。本論文では、数値推論タスクにおいて、意味的に類似した例を区別しながら、異なる数式を維持することを可能にする対照学習の枠組みを提案した。自己教師ありの戦略により、テキストの微妙な変化に基づき、数式とテキストの両方の観点から最も困難なサンプルを捉え、その表現を学習するようモデルを導く。さらに、加法交換則、乗法交換則など5つの制御コードが指定可能な数式生成器を設計し、数値推論におけるモデル学習を頑健化した。数学文章問題においてこれらの手法の有効性を確認した。
3. 数値推論タスクの弱教師付き学習における重要な課題として、膨大な探索空間とアライメント誤りの問題に取り組んだ。数学的に等価な数式をまとめるための組み合わせ戦略を取り入れた新しい探索アルゴリズムを考案し、このアルゴリズムによって探索空間を圧縮し高品質なデータを取得することを可能とした。また、誤った数式を持つ擬似ラベルによるノイズの調査を行い、そのようなインスタンスを除去することでノイズを軽減するランキングモデルを提案した。これにより標準的な数学推論データセット Math23K において 4.9 ポイントの精度改善を実現した。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、令和5年8月28日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。