

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	2311229	提出日	令和 7年 1月 21日
学生氏名	日浦 隆博		
論文題目	敵対的学習を用いた記号的知識蒸留		
要旨			
<p>大規模言語モデル(LLM)は多くの自然言語処理タスクで顕著な性能を実現しているが、知識推論のようなタスクにおいては依然として出力する知識の精度に課題が残っている。これに対して、知識グラフの部分グラフであるトリプルを用いてLLMをファインチューニングし、知識間の関係を明示的に学習させることで、推論精度の向上を図るアプローチが注目されている。これには、大規模で高精度な学習用トリプルデータセットの獲得が重要であり、有効な手段として記号的知識蒸留がある。これはLLMの持つ膨大な知識をトリプルとしてサンプリングし、知識推論モデルの学習に使用する手法であり、既存のトリプルデータセットの大規模化が期待できる。しかしLLMからサンプリングしたデータには意味的・文法的に誤った出力が一定数含まれており、学習データの品質に影響を与えるため、それらを除去するフィルタが必要となる。本研究では、フィルタリング(選択)を行うモデルと、選択されたデータを評価するモデルを敵対的に学習することで、LLMの出力を自動的にフィルタし、トリプルデータセットを大規模化する枠組みを提案する。こうして獲得したフィルタは、比較手法と比べてprecisionが向上するという結果になった。またフォーマットに問題のあるデータなどをより適切に識別できるようになることも分かった。</p>			