

Retrieval and Disambiguation of Mathematical Expressions for Mathematical Information Access

その他のタイトル	数学情報アクセスのための数式表現の検索と曖昧性解消
学位授与年月日	2017-03-23
URL	http://doi.org/10.15083/00076183

審査の結果の要旨

氏 名 クリステリアント ギオヴァニ ヨコ

本論文は、科学技術関連情報を扱う上で欠かすことのできない数式情報の検索について、自然言語情報を併用することで精度を向上させることができることを、体系的に論じたものである。これまで数式情報の検索に関してはその重要性からいくつかの研究があったが、複雑な構造を持つ数式を計算機で検索することの困難性、関連情報との関連付けの困難性、同一の概念を様々な数式で表現可能であることから来る曖昧性などの様々な問題から、精度の高い検索は極めて困難であった。これに対し本論文では、そのような数式情報検索の精度を高めるための、二つのこれまでにない全く新しい検索戦略を提案している。一つは数式情報とそれと関連する自然言語情報を効果的に構造化し、それを索引として持つことでより高度な検索を実現する戦略である。もう一つは、数式情報間の関連を抽出しそれを活用することで自然言語情報をより濃縮された情報とするエンリッチメント戦略である。さらに本論文では、これらの二つの戦略に基づくことで実際にも高精度な検索が実現可能であることを計算機実験を通して実証することに成功している。また、この他、これらの数式情報検索技術の応用として、数式のウィキフィケーションなどエンティティ・リンキングへの応用についても扱い、その効果を検証している。

本論文は七章からなり、第一章では、本論文の背景と動機として、数式情報検索の課題としてどのような課題があるのか、そのフレームワークについて論じ、本論文全体の貢献の重要性について概観している。

第二章では、本論文における議論の準備として、実際の数式データの特徴、数式情報を扱った既存の手法、エンティティ・リンキングや非曖昧化の手法について網羅的な調査を行っている。

第三章では、数式を効果的に検索することを可能とするための数式構造情報の計算機上での表現方法、数式と関連した自然言語の抽出方法およびその表現方法について論じ、さらにそれらの索引方法、検索方法を提案している。さらにそれらの手法について、実際に数式を含む文書に対して計算機実験を行い、その精度の検証を行っている。

第四章では、数式情報検索の精度をさらに上げることを可能とするために、数式間の依存関係を表すグラフを抽出する手法を提案している。また、抽出されたグラフ情報を

活用し、より濃縮された自然言語情報を抽出するエンリッチメント手法を提案している。さらに、その濃縮された情報を活用することで数式情報検索の精度を高めることが可能であることを、実際の文書に対する計算機実験を通して評価検証している。

第五章ではさらに、第四章において提案したエンリッチメント手法の応用として、ウィキフィケーション等のエンティティ・リンクングを数式に対して行う技法を提案するとともに、実際の文書を用いた計算機実験を通じその性能を評価検証している。

第六章では、第五章までの手法や問題について総括するとともに、将来の展望に関して議論を行っている。

第七章では、本論文の貢献について要約し、まとめている。

このように本論文は、科学技術情報検索において重要な数式検索という分野に対して、これまでなかった数式情報と自然言語情報を構造的に統合するまったく新しいアプローチを導入することによって既存の手法などと比較してきわめて高度・高精度な検索を実現したものであり、これまでにない新しい研究分野を切り拓くとともに、当該分野の今後の発展にも大きくつながることが期待できる。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。