

## 研究課題別事後評価結果

1. 研究課題名： 数式と自然言語の統合的解析による学術文献理解の研究
2. 個人研究者名  
朝倉 卓人（東京大学大学院情報理工学系研究科 大学院生）
3. 事後評価結果

本研究では、数式グラウンディング、すなわち数学論文中に記述された数式記号（例えば、 $x$  や  $y$ ）がどのような数学概念に対応しているかを同定するという、極めて挑戦的な課題に取り組んだ。その課題の第一歩として、15 編の論文について、その中の記号（総数 1.2 万）とその数学概念を対応付けたデータセットを構築した。その結果、本来曖昧性のない明確な記述が必要とされる論文において、多くの記号が 2 つ以上の解釈が可能（例えば論文の途中で意味が変わる）という、驚くべき結果を定量的に得ている。現時点では、こうした課題の困難性の把握、データセット構築環境 (Mio Gatto)、およびデータセット記述フォーマットを、海外の研究者とのディスカッションも行いながら実現した段階である。今後はこれら知見を活かし、自動対応付け (アノテーション) や、意味の切り替わり (スコープ) の自動判定を実施予定とのことである。同技術は、知の集積としての論文に記述された内容を二次利用するためには不可避であり、その嚆矢として、本研究の今後の発展が期待される。