

## 階層関係に基づく数学教材検索のための階層的類似メトリック方式

笠井 裕貴<sup>†</sup> 中西 崇文<sup>†</sup> 岡田 龍太郎<sup>†</sup>武蔵野大学 データサイエンス学部 データサイエンス学科<sup>†</sup>

## 1. はじめに

一般に、高校生を対象とした数学の学習者が数学の問題を解く際に、問題文からどの公式を使うのかをある程度予測する能力を養うことが重要である。数学の学習者を対象として、学習者が入力した問題文に対し、その問題を解くために必要な公式とその公式を使うに至った根拠を出力し、さらに入力された問題と類似した問題を出力するシステムが実現できれば、学習者にとって解けなかった問題についての知識を得ることができるだけでなく、類似問題を得ることで、その問題を解き、数学の学力を着実に身につけることができると考えられる。

本稿では、階層関係に基づく数学教材コンテンツ検索のための階層的類似性メトリックの実現方式について示す。本方式は、数学の問題文及び数式を入力とし、その問題が属している単元を特定し、その単元であると推定に至った根拠を出力する。本方式は、事前に準備されている問題文データセットから各単元を表す単語群を抽出し、入力された問題文に出現する単語群で類似度計量を行うことで、どの単元の問題かを推定する。また、本方式では、この類似度計量を階層的に行う。階層的に行うことによって、一般的に数学科目において、単元が階層的に構成されていることから、学習者にとって単元とその公式を結びつけて記憶する助けとなりうる。本方式は、上位階層から順に類似度計量をすることによって問題文同士の類似性を計量することを階層的類似性メトリックと定義し、数学科目の単元の階層性を用いた階層的類似性メトリックを実現する。さらに、本方式では、推定した単元の根拠を示すため、単元固有の単語が入力した問題文中の単語と合致した場合に強調して表現することを可能としている。

## 2. 関連研究

狩山ら[1]は、文字列と数式の混ざった数式問題

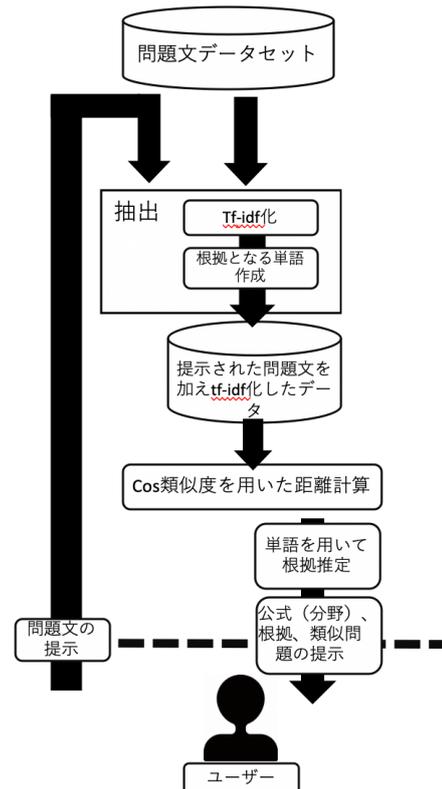


図1：階層的類似性メトリック方式の全体像

文の検索を実現している。我々の提案方式では、数学科目の単元から構成されている階層構造を利用し、類似性を絞り込む階層的類似性メトリックを実現することにより、類似問題を検索することを可能とする。

## 3. 提案方式

## 3.1 全体像

図1に階層的類似性メトリック方式の概要を示す。本方式は大きく分けて、単語抽出部、類似度計量部、出力部によって構成される。

## 3.2 単語抽出部

本方式では、根拠を提示するための単語を生成するために、問題文データセットを単元ごとに分け、単元を1つの文書と見立ててTf-idf値を求め、単元ごとに数値の大きかった単語をその単元を特定するための単語群として設定する。

階層的分析の例

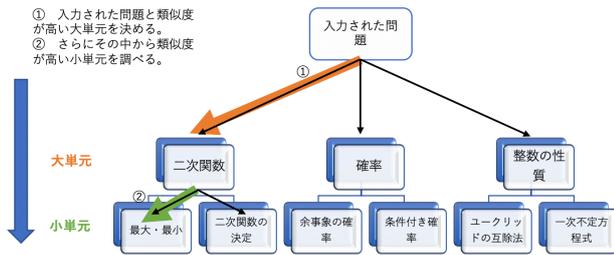


図 2：階層的分析の例

3.3 階層的類似性メトリックによる類似度計量部

図 2 に本方式で実現する階層的類似性メトリックの概要を示す。数学問題の単元の階層に基づく類似度計量を行うために、単元ごとに分けられた問題文データセットの単語群と入力された問題文の単語群との類似度を求めて、どの単元の問題かを推定する。

また、図 2 のように、階層的分析を行うため、最初に二次関数、確率などの大単元の推定を行い、さらに、下層の小単元の推定を行う。つまり、問題文データセットを様々な大単元、小単元に分け、それぞれ Tf-idf 値を求めることで、各層において、各単元の単語群を導出し、入力された問題文との類似度を求めることが可能となる。

3.4 出力部

出力部では、3.2 節で導出された大単元、小単元、およびその根拠が導出される。ここで根拠とは、入力された問題文が推定された大単元、小単元であると決定づける単語である。各単元に設定した単語群が、入力された問題文と一致する部分があればその単語を強調表示する。

4. 実装

4.1 データセット

図 1 内の問題文データセットとして、「教科書より詳しい高校数学」[2]からスクレイピングして、データセットを作成した。

4.2 結果出力例

本稿では「関数、 $ax^2-2ax+b(-2 \leq x \leq 2)$  の最大値が 5 で、最小値が -4 であるとき、定数 a, b の値を求めよ」という問題文を入力としたときの出力結果について示す。

表 1 に入力された問題文と大単元との類似度の結果を示す。その結果、「二次関数」単元の類似度が高くなっていることが分かる。さらに、「二次関数」単元の下に属する小単元との類似度結果を表 2 に示す。「最大値・最小値 2」単元の類似度が高くなっていることが分かる。この入力された問題文の単元を正しく推定できてい

表 1：大単元における類似度計量結果

コサイン類似度	
二次関数	0.471398
図形と計量	0.285327
データの分析	0.247904
数と式	0.076708
集合と論理	0.010013

表 2：小単元における類似度計量結果

コサイン類似度	
関数の値	0.467672
二次関数のグラフ	0.082065
グラフの移動	0.000000
二次関数の決定	0.563951
最大値・最小値	0.378006
最大値・最小値2	0.860269
二次方程式の解	0.117076
解が与えられた二次方程式	0.104970
二次方程式の文章問題	0.006829
二次関数と交点	0.153523

関数、 $ax^2-2ax+b(-2 \leq x \leq 2)$  の最大値が 5 で、最小値が -4 であるとき、定数 a, b の値を求めよ

図 3: 根拠となる単語を斜体として表現

ることが確認できる。

図 3 に示すように、単元の推定結果について、問題文のどの単語が根拠となっているかを斜体で表現している。

5. おわりに

本稿では、階層関係に基づく数学教材コンテンツ検索のための階層的類似メトリックの実現方式について示した。本方式は、上位階層から順に類似度計量をすることによって問題文同士の類似性を計量することを階層的類似性メトリックと定義し、数学科目の単元の階層性を用いた階層的類似性メトリックを実現する。さらに、本方式では、推定した単元の根拠を示すため、単元固有の単語が入力した問題文中の単語と合致した場合に強調して表現することを可能としている。

今後の課題として、本方式を応用した類似問題の提示手法の実現、他の科目の分野への方式の適用が挙げられる。

参考文献

[1] 狩山和亮, 吉仲亮, 山本章博, 文字列パターンと MathML による構造を利用した数学問題文の検索, 人工知能学会全国大会論文集 第 27 回全国大会, 1F51, 2013.  
 [2] <https://yorikuwa.com/>