

1 はじめに

大学でどのような研究を行うかということはその後の人生に多大な影響を与える要素である。そのため、高校生が進学先を選択する際、各大学及び研究室で過去にどのような研究が行われていたかを知ることが重要である。しかし、今日では研究テーマが多様化し、これらを調べることは手間がかかるため、特に受験生にとって大きな負荷となってしまう。そこで本研究では、大学で学生が行った研究論文のタイトルデータ群に対し、テキストマイニングの手法を用いることで、研究テーマを表や図、グラフの形で可視化し、受験生にわかりやすく提示することを目的とする。同時に、研究テーマの年度ごとの傾向についても分析を行う。

2 研究室ごとのテーマの可視化・分析手法

2.1 用いるデータ

2008 年度から 2017 年度の名古屋工業大学情報工/メディア系の学部生及び大学院生が行った研究論文のタイトルをデータとして用いる。各論文データは、論文タイトル、年度、研究室、卒業論文または修士論文の情報を持つ。

2.2 語の抽出

まず各論文タイトルデータを、テキストマイニングの技術を用いて語に分解する。抽出された語のなかで、出現回数が多いものをまとめ、考察を行う。また、抽出された語のなかで、研究テーマに関連性の薄い「用いる」などの語は、あらかじめ分析対象から除外しておく。また、抽出した語同士の共起関係を表した共起ネットワークを作成する。さらに、論文データが持つ研究室、年度、卒業論文または修士論文の 3 つの要素と、抽出された語との関連性についてそれぞれ分析を行う。そのために、各要素の内容と抽出された語との関係性をそれぞれネットワーク図で表す。

2.3 カテゴリ分類

研究テーマの傾向をさらに簡単に理解しやすくするため、語の抽出を行った結果から、頻出する語のなかで同義語や類義語を見つけ出し、それらの語をまとめてカテゴリを作成する。その後、カテゴリに設定した語がタイトル中に含まれる論文データをそのカテゴリに分類し、カテゴリに含まれる論文の数をもとに分析を行う。

3 分析結果

名古屋工業大学のタイトルデータを語に分解した結果、出現回数が多かった上位 10 個の語を表 1 に示す。この表から音声分野及び画像分野に関連する語が多く抽出していることがわかる。語と研究室との関連性

の分析では、各研究室の特徴について把握することができた。例として、語と研究室との関連性を表したネットワーク図のうち、舟橋研究室に関する部分を抜粋したものを図 1 に示す。この図より、舟橋研究室は「VR」や「仮想」など VR に関する研究と関連性が強いことがわかる。また、カテゴリ作成では「音声」「画像」「機械学習」などの計 7 つのカテゴリを作成し、分析を行った。その結果、各研究室のおおまかな研究テーマを可視化することができた。また、「機械学習」カテゴリに対して、「ディープラーニング」などの機械学習の手法に関するカテゴリを新たに作成し、機械学習手法の年度推移についても分析を行い、傾向を見出した。

表 1: 語の出現回数

	抽出語	出現回数		抽出語	出現回数
1	音声	165	6	HMM	62
2	モデル	115	7	情報	62
3	認識	104	8	生成	61
4	画像	102	9	学習	55
5	対話	78	10	合成	55

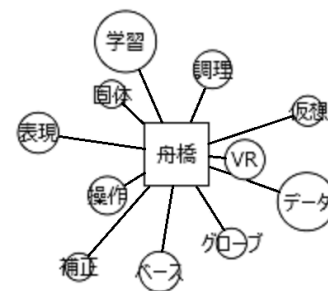


図 1: 研究室傾向 (舟橋研)

4 むすび

本研究では、学生の研究論文のタイトルに対して、テキストマイニングの技術を用いることで、大学研究室で行われている研究のテーマの可視化と分析を行った。その結果、名古屋工業大学の各研究室の研究テーマを作成した図から捉えることができた。また、年度ごとの研究テーマの傾向についても分析・考察を行った。さらに、東京大学のデータに対しても同様の分析を行い、各研究室の特徴を掴むことができています。今後の課題としては、年度推移の分析による今後の研究テーマのトレンドの予測や、自動的にカテゴリを作成するシステム的设计などが挙げられる。将来的には、ひと目で各研究室、学科、大学ごとの研究テーマが理解できるような可視化を行い、受験生の進路を決める指針となることを期待している。