平成30年度 卒業論文

テキストマイニングによる
大学研究室ごとの研究テーマの可視化と分析

指導教員
舟橋　健司　准教授
伊藤　宏隆　　助教

名古屋工業大学　工学部　情報工学科
平成27年度入学　27115110番

名前　長縄　龍風

目 次

第1章　はじめに

近年，SNS などの IT 技術の普及と発達により，世の中には大量のデータが蓄積されている。そのため，これらを分析し有用な情報や傾向を見つけ出すデータマイニングが注目を集めており，医療や商業の分野で利用されている。例として、商品の月や時間ごとの売上データから，顧客の需要の傾向を発見し供給量を調整することで利益の向上が図られていることや，医療分野での臨床データに基づいた病気の経過予測などが挙げられる．

教育現場でも，データマイニングの技術を用いて学生の成績などから一人ひとりの修学傾向を読み取り，学習指導を行うという提案がされている．過去の関連研究では，各学生の成績と授業アンケートから，成績と学生の行動の関係，成績と授業の進行の仕方との関係などを調査したもの [1] や，留年や退学をする学生を調査・分析し，事前にこれらの学生を予測する研究 [2] などがある．

また，データマイニングの中でも文字列を分析対象としたものはテキストマイニングと呼ばれる．テキストマイニングの技術を用いた研究では，講義に対しての意識調査のアンケート結果を分析し，指導内容への受講生の提言をまとめたもの [3] や，SF 作品の文章から傾向を見出し，未来の社会の有り様を予測したもの [4] などがある．

ところで，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．例えば就職活動でも大学で行った研究のテーマが雇用されるか否かのひとつの指標となりうる．そのため，高校生が進学先を選択する際，各大学，学科，研究室で過去にどのような研究が行われているかを調べることは重要である．しかし，今日では研究テーマも多様化し，論文数も膨大であるため，これらを調べることは手間がかかり効率的でない．特に受験生にとって，このことは大きな負荷となってしまう．

そこで，本研究では大学の研究室で行われている研究のテーマの分布や傾向を表や図，グラフなどの形で可視化することで，情報を視覚から安易に入手できるようにする．その手法として，卒業論文，修士論文のタイトルデータ群に対してテキストマイニングの技術を用いることで，研究テーマに関する単語を表，図，グラフ化し，学科全体及び研究室ごとの研究テーマの傾向を見出す．さらに，得られた結果からいくつかのカテゴリを作成し，各研究を作成したカテゴリに分類する．その後，各カテゴリと各研究室との関係を図示することで，研究テーマの傾向をより視覚的にわかりやすくする．また，並行して年度ごとの研究テーマの傾向についても同様に分析することで，研究の推移や流行についても考察する．

本論文では，まず第 2 章で本研究で用いるテキストマイニング技術の理論について，第 3 章では分析で用いるデータの概要とどのような分析を行うかの説明，第 4 章で行った分析の結果と考察，最後に第 5 章で本研究のまとめと今後の課題や展望を述べる．

第2章　テキストマイニングの理論

本研究では，テキストマイニングの技術を用いて多くの分析を行っている．本章ではその技術について説明する．

2.1　テキストマイニングの概要

テキストマイニングとは，テキスト型データの分析手法のひとつである．文章データを単語や文節で区切り，それらの出現回数や出現傾向などをもとに解析を行う．そうして文章の特徴を可視化することで，短時間で有用な情報を得ることが可能となる．テキストマイニングによる分析を行うことで，文章のおおまかな全体像を把握することができたり，特定の単語に関する特徴を抽出することができる [6].

2.2　テキストマイニングによる分析の手順

テキストマイニングによる分析は，次のような手順で行われる．

1. 分析対象となるデータを用意し，分析できるよう形を整える．
2. 文を単語単位に分解する．
3. 単語の出現回数などの情報をもとに表，図，グラフを作成する．
4. 作成した表，図，グラフを分析，考察する．

手順2について，2.3節で詳しく説明する．また，手順3で作成する図のひとつである共起ネットワークについて2.4節で詳しく説明する．

2.3　文の分解

文を単語単位に分解するために，形態素解析を行う．形態素解析とは，文法や単語の品詞などの情報にもとづき，文を意味を持つ最小単位（形態素）に分解、各形態素の品詞などを判断する手法である．日本語に対する形態素解析では，語の境界を判別することが困難であるため，近年では統計的な手法が多く用いられている [5].

形態素解析の例として,「吾輩は猫である」という文に対して形態素解析を行った場合,「吾輩/は/猫/ で/ある/」のように分解され，各形態素にそれぞれ「代名詞/助詞/名詞/助動詞/動詞」といった品詞が割り当てられる．

また，本研究では分析の精度向上のため，動詞や形容詞などの活用形のある語はすべて基本形に直して抽出される．例えば,「買え」という語は「買う」の活用形であるため,「買え」と「買う」という語がデータ中にひとつずつ存在した場合,「買う」がふたつ出現したものとみなされる．

2.4　共起ネットワーク

ある単語とある単語が同時に文中に出現することを共起という．共起関係にある単語どうしを線で結び，ネットワーク図として表したものが共起ネットワークである．共起ネットワークの例として，夏目漱石の「こころ」の分析結果を図 2.1 に示す．ただし，円の大きさは語の出現回数，線の濃さは共起の強弱，色分けはお互いに強く結びついている語のグループを表す．また，各ノード間の距離は意味を持たない．図 2.1 から，例えば「K」「奥さん」「お嬢さん」という語の結びつきより，これらの人物は同じ場面に登場しやすいことなどがわかる．

共起関係をすべて描くと画面が埋まってしまうため，描画する関係を絞る必要がある．本研究では，共起の程度を値で算出し，その値に基づき描画する共起関係を選択する．この値を算出するために，Jaccard 係数を利用する．Jaccard 係数とは，集合同士の類似度を測定する際，主に用いられる手法のひとつである．ある単語 A と B の共起の程度を測定する場合，語 A を含む文書の集合を S_A，語 B を含む文書

の集合を S_B とすると，算出する Jaccard 係数 $J(S_A, S_B)$ は以下の式で求められる．

$$J(S_A, S_B) = \frac{|S_A \cap S_B|}{|S_A \cup S_B|} \tag{2.1}$$

算出された Jasscard 係数が大きいほど共起の程度が強い．

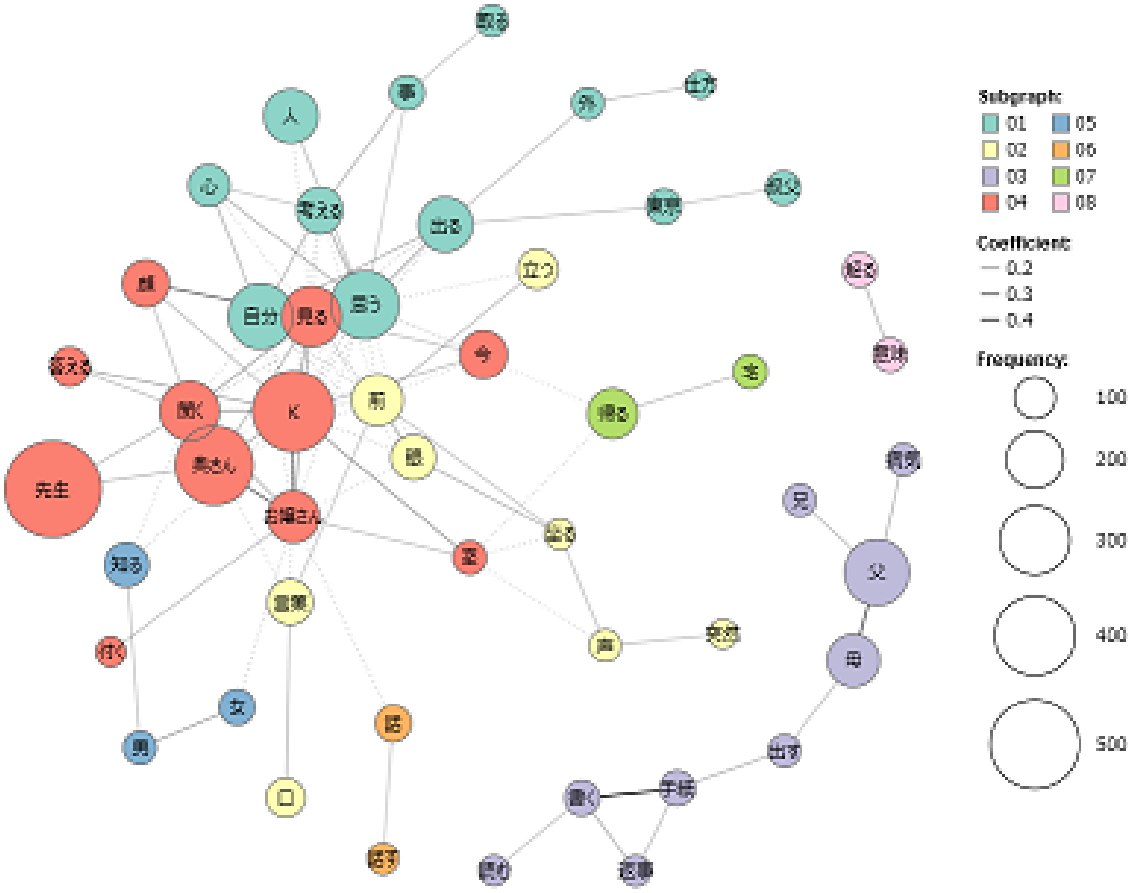

図 2.1: 共起ネットワークの例

第3章　データの概要と分析の手法

本章では，本研究に用いるデータの概要と，分析の手法について述べる．

3.1　データの概要

本研究では，大学で学生が行った研究論文（卒業論文及び修士論文）のタイトルをデータとして用いる．今回は，名古屋工業大学の 2008 年度から 2017 年度の 10 年分のデータを用いる．さらに，このデータとの比較のため，工学部のなかで同年度分が Web 上 [7] で公開されている東京大学のデータを用いる．本節では，これら 2 種類のデータについて述べる．

3.1.1　名古屋工業大学のデータ

名古屋工業大学のデータは，情報工学科メディア系に所属していた学生の卒業論文のタイトルと、大学院情報工学専攻メディア情報分野に所属していた学生の修士論文のタイトルをまとめたものである．ひとつの論文データは，論文のタイトル，論文が発表された年度，論文を発表した学生が所属する研究室，論文が卒業論文であるか修士論文であるかの 4 つの情報を持つ．年度ごとの論文データ数を表 3.1 に示す．また，各研究室ごとのデータ数を表 3.2 に示す．

3.1.2　東京大学のデータ

東京大学のデータは，工学部機械情報工学科に所属していた学生の卒業論文のタイトルデータである．ひとつの論文データは，論文のタイトル，論文が発表された年度，論文を発表した学生が所属する研究室の 3 つの情報を持つ．なお，東京大学のデータは卒業論文のみであるため，卒業論文であるか修士論文であるかの情報は

持たない．年度ごとの論文データ数を表 3.3 に示す．また，各研究室ごとのデータ数を表 3.4 に示す．

表 3.1: 論文データ数 (名工大)

年度	卒業論文	修士論文
2008	63	22
2009	53	28
2010	57	40
2011	54	34
2012	51	32
2013	56	37
2014	58	33
2015	56	42
2016	46	28
2017	51	54

表 3.2: 各研究室のデータ数 (名工大)

研究室	論文数
徳田	158
佐藤	113
岩田	96
本谷	85
舟橋	70
黒柳	69
李	65
北村	62
梅崎	46
山本	38
白岩	38
小田	30
松添	11
平野	9
夏目	4
クグレ	1

表 3.3: 論文データ数 (東大)

年度	論文数
2008	43
2009	37
2010	43
2011	43
2012	44
2013	41
2014	41
2015	46
2016	41
2017	50

表 3.4: 各研究室のデータ数 (東大)

研究室	論文数
廣瀬	69
國吉	68
稲葉	65
中村	55
下山	49
神埼	36
佐藤	30
原田	29
土肥	17
下坂	6
正宗	4

3.2　分析の手法

各データに対して，テキストマイニングの技術を用いて分析を行う．本研究の分析には，フリーソフトウェアである「KH Coder」[8] を使用した．

3.2.1　語の抽出

論文データの「論文のタイトル」の項目を対象に，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語をまとめ，表にする．なお，本研究において出現回数とは「出現した論文の数」を意味する．例えば,「A を用いた A システム」というように，ひとつのタイトル中に A という語が 2 度出現したとしても，A の出現回数は 1 論文として扱う．

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．同時に視覚的により多くの情報を得られるよう，共起の強弱を線の濃さで表し，語の出現回数を語のノードの大きさで表す．さらに，分析の参考として，比較的強くお互いに結びついている部分を自動的に検出してグループ分けを行い，その結果を色分けによって示す「サブグラフ検出」を行う．

3.2.2　語の取捨選択

抽出される語のなかで，助動詞，助詞，接続詞，連体詞は一般的に具体的な意味を持たないことが多い．さらに，例えば「用いる」のように，多くの論文に含まれており、なおかつ一般的に研究テーマに関する情報を持たない語が存在する．このような語は，結果を可視化した際，観測者の注意をひきつけてしまい，分析の妨げとなる．そのため，これらの語に関しても分析の対象外として設定する．対象外とする語は以下の通りである．

用いる，基づく，研究，手法，開発，システム，実装，実現，検討，検証，する

また，語を抽出する際,「スマートフォン」のようにひとつの語として抽出すべき語（複合語）が,「スマート」と「フォン」といったように 2 つ以上の語として分割されて抽出されてしまう場合がある．このように抽出されてしまうと分析を正しく

行うことができないため，あらかじめひとつの語として抽出するよう設定する．このように設定する語は以下の通りである．

スマートフォン，ライトフィールド，レジストレーション，インタラクション，データベース，モーダル

3.2.3　語と外部変数

論文データが持つ 3 つの要素「研究室」「年度」「卒業論文または修士論文」を外部変数として，論文タイトルから抽出された語との関連性について分析を行う．まず，語と各研究室との関連性について分析を行う．そのために，各研究室で発表された論文のタイトルのなかで頻出している語のノードを，その研究室のノードと線で結んだネットワーク図で表す．なお，語と各研究室との関連性の強弱は共起ネットワークと同様 Jaccard 係数を使用する．また，視覚的に理解しやすくするため，語が結びついている研究室の数によって色分けを行う．

また，語と各年度との関係，語と修士論文であるか卒業論文であるかの関係についても同様にネットワーク図で表す．

3.2.4　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，論文データをいくつかのカテゴリに分類し，分析を行う．語の抽出を行った結果から，頻出する語のなかで，研究テーマが似ているものを見つけ出し，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に含まれる論文データをそのカテゴリに分類する．例えば,「音」「音響」といった音声に関する語が多く出現していた場合,「音声」というカテゴリを作成し,「音」や「音響」などの語をタイトルに含む論文データを,「音声」カテゴリに分類する．なお，ひとつの論文データが複数のカテゴリに分類される場合もある．また，どのカテゴリにも属さない論文データも存在する．

このように，カテゴリ分類を行ったのち，カテゴリに含まれる論文の数をもとに，各カテゴリの傾向，および各研究室とカテゴリとの関係，カテゴリの年度ごとの傾向を表，図，グラフで表し，分析する．

3.2.5　データの比較

3.2.1 項から 3.2.4 項までの分析を，名古屋工業大学のデータに対して行う．その後，同様の方法で東京大学のデータに対しても分析を行い，分析の有用性を確かめると同時に，2 つの大学のデータを比較し考察する．

第4章　分析結果

本章では，名古屋工業大学と東京大学のデータに対し，3.2 章で解説した分析を行った結果と考察を述べる．

4.1　名古屋工業大学のデータを用いた分析

本節では，名古屋工業大学のデータに対して行った分析の結果と，それに対する考察を述べる．

4.1.1　抽出された語

3.2.1 項で述べたように，論文タイトルから語の抽出を行った．表 4.1 に，抽出された語の中で出現回数が多いもの上位 10 個を示す．また，この抽出結果をもとに共起ネットワークを作成した．この共起ネットワークを図 4.1 に示す．ただし，この図は程度の強い共起関係上位 100 個を表しており，また 10 回以上出現していない語はノードに含まれない．

表 4.1: 取捨選択前の語の抽出結果

抽出語	出現回数
研究	222
用いる	209
システム	183
基づく	178
音声	165
する	165
モデル	115
認識	104
画像	102

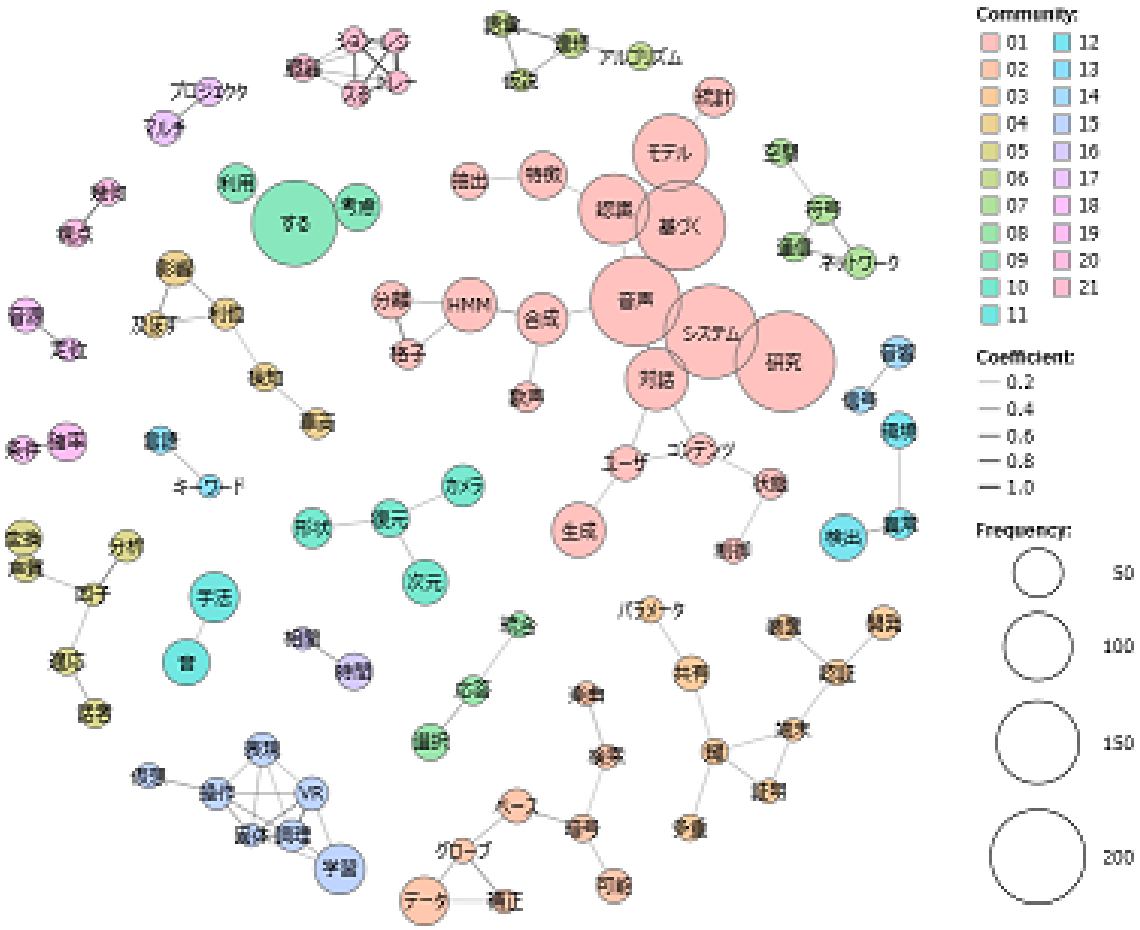

図 4.1: 語の取捨選択前の共起ネットワーク

この抽出結果から,「用いる」や「システム」,「基づく」などの研究テーマに関連性の薄い語が多く抽出されていることがわかる．これでは，本来確認したい語が下位に埋もれてしまい，分析が困難であることがわかる．また，共起ネットワークでも「基づく」「する」「研究」などの不要な語が図の多くを占め，観測者の注意を引いてしまう．したがって，これらの不要な語を分析から除外する必要がある．

4.1.2　語の取捨選択後

3.2.2 項で述べた語の取捨選択を行い，再度語の抽出を行った．表 4.2 に，抽出された語の中で出現回数が多いもの上位 20 個を示す．また, この抽出結果をもとに共起ネットワークを作成した．この共起ネットワークを図 4.2 に示す．ただし，この図は程度の強い共起関係上位 100 個を表しており，また 10 回以上出現していない語はノードに含まれない．

表 4.2: 語の抽出結果 (名工大)

抽出語	出現回数
音声	165
モデル	115
認識	104
画像	102
対話	78
HMM	62
情報	62
生成	61
学習	55
合成	55
推定	51
自動	49
特徴	46
考慮	45
データ	44
検出	44
音	41
次元	40
カメラ	37

語の抽出結果より，まず表 4.1 で存在した「研究」や「用いる」などの語が除外され，必要な情報を得やすくなっていることがわかる．次に,「音声」「画像」の 2 つの語が極めて多く抽出されている．このことから音声分野と画像分野の研究が主として行われているといえる．加えて,「音声」は「画像」よりも出現回数が多く,「対話」「音」といった語も多く出現していることから，音声分野の研究が特にさかんであるといえる．また,「HMM」や「学習」といった語も多く出現しているため，機

械学習を用いた音声や画像に対するアプローチが行われていると推測できる．

また，共起ネットワークを分析する．まず，語の抽出結果と同様に，図 4.1 では存在していた「研究」「する」などの語が除外され，分析しやすくなっていることがわかる．次に,「音声」という語は,「対話」「合成」「認識」との共起が強く，音声を扱う研究では主に音声対話，音声合成，音声認識が研究されていることがわかる．また,「合成」は「HMM」と極めて強く結びついているため，HMM を用いた音声合成がさかんに行われてきたことが伺える．一方で,「画像」は「分離」「格子」と共起関係があるが,「画像」の出現数に対して「分離」「格子」の出現数が少なく，共起の程度も弱い．これは，画像を扱う研究は多岐にわたるため，共起の強い語が現れなかったと推測できる．

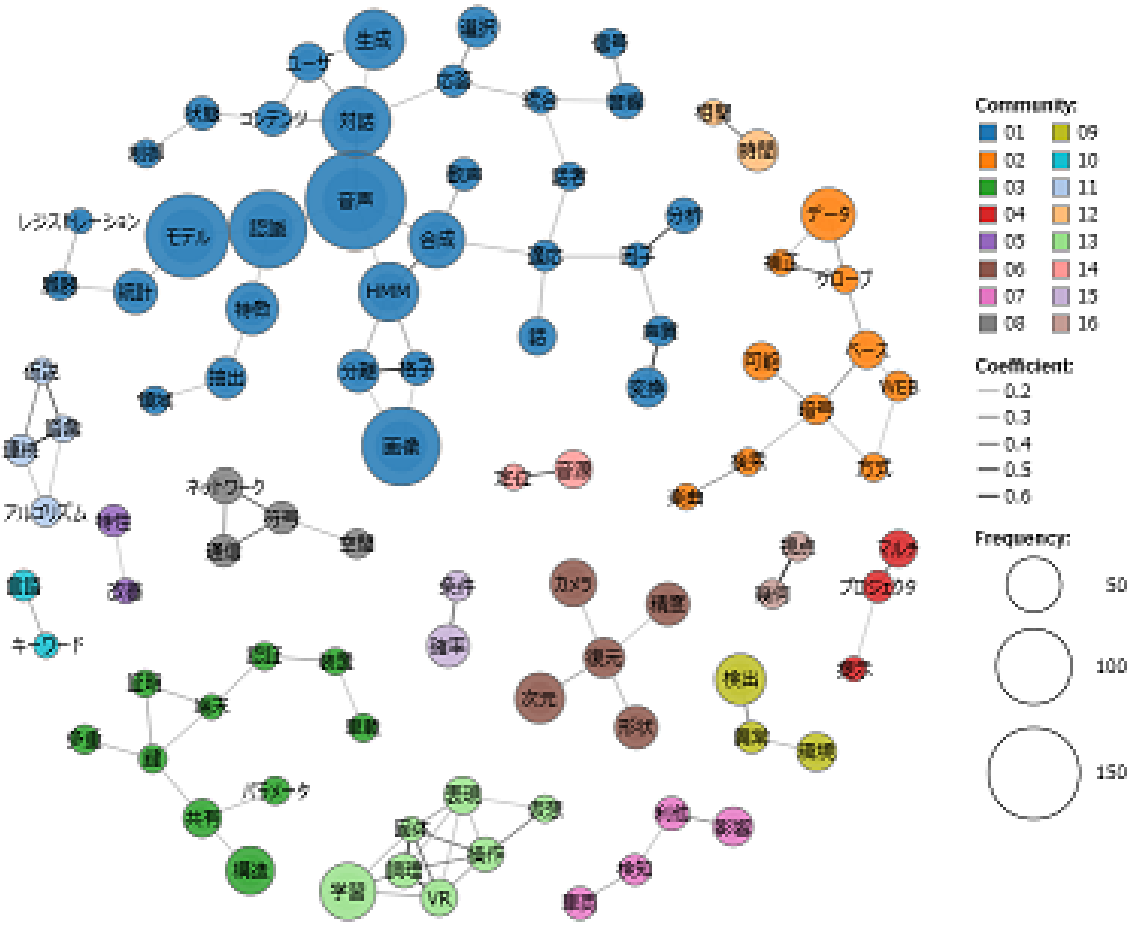

図 4.2: 共起ネットワーク (名工大)

4.1.3　語と研究室の関連性

3.2.3 項で述べたように，語と各研究室との関連性について分析を行った．語と研究室との関連性を表したネットワーク図を図 4.3 に示す．ただし，この図は程度の強い関係上位 100 個を表しており，また 10 回以上出現していない語はノードに含まれない．

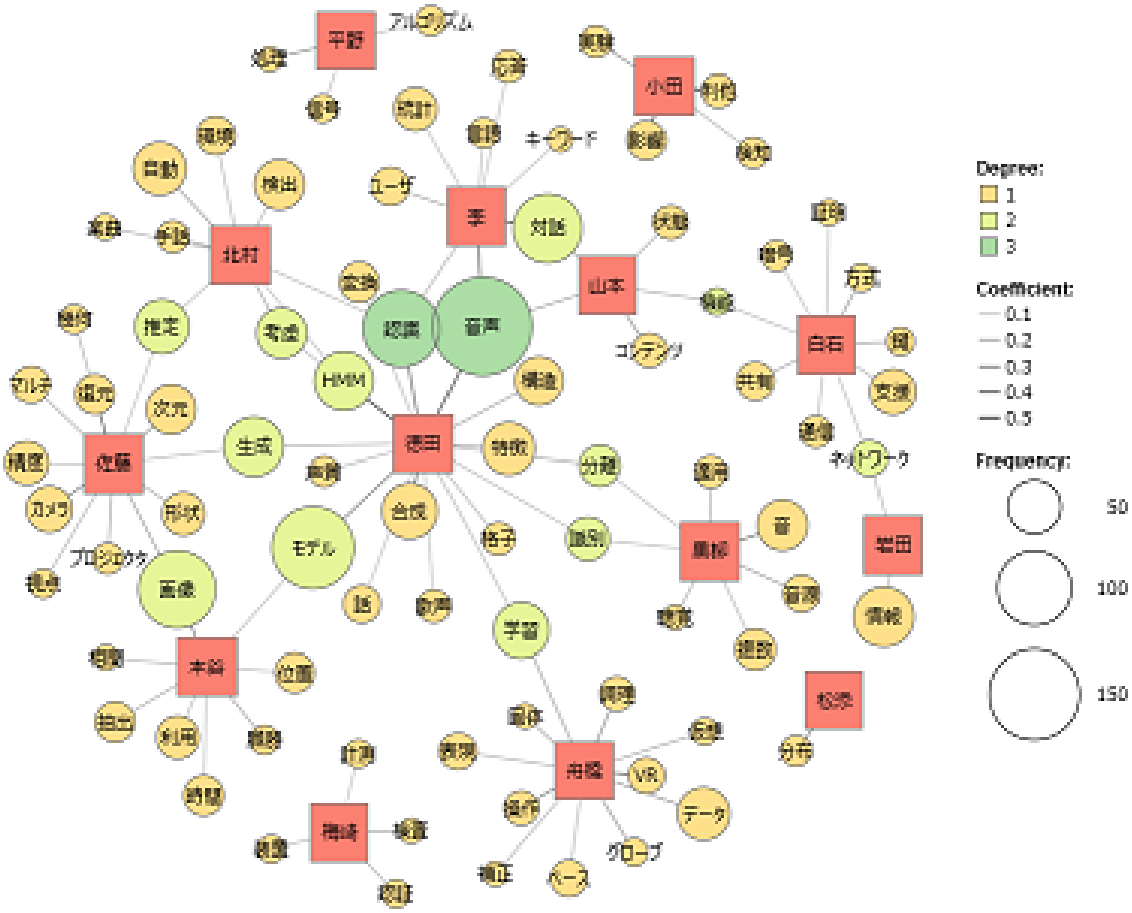

図 4.3: 研究室傾向 (名工大)

この図から，各研究室について以下のようなことが読み取れる．

徳田研：「音声」との関連性が強くでていることより，音声分野の研究が主であることがわかる.「認識」「合成」とのつながりもあることから，なかでも音声認識及び音声合成に関する研究が行われているとわかる．また,「HMM」との関連性が極めて強いため，前項の分析も考慮すると，この研究室が主体となって

HMM による音声合成の研究を行っていたと推測できる.

李研： 「音声」との関連性が強くでていることより，音声分野の研究を主であることがわかる.「対話」「認識」とのつながりがあるが,「対話」のほうが関連性が強く,「言語」「キーワード」「応答」といった語ともつながりがあることから，音声対話に関する研究がよりさかんであると推測できる.

山本研： 「音声」とのつながりがあることから，音声分野の研究が主であるとわかる.「対話」とのつながりもあることから，音声対話に関する研究が行われているとわかる．また,「コンテンツ」「機能」といった語とつながりがあるため，音声対話を扱うコンテンツの制作を行っていると推測できる.

北村研： 「音声」とのつながりは見られないが,「楽曲」とつながりがあることから音声分野の研究が行われていることがわかる．一方で,「手話」とのつながりもみられる．手話は主に聴覚障害者のためのものであるから，これは画像関連の研究であると推測できる．このことから画像分野と音声分野の両方を研究していると考えられる.

黒柳研： 「音声」とのつながりは見られないが,「音」「音源」「聴覚」とのつながりがあることから，音声分野の研究が主であるとわかる．また,「識別」「分離」とのつながりから，音の識別や分離に関する研究が行われていると推測できる.

佐藤研： 「画像」とのつながりがあり,「カメラ」「視点」「プロジェクタ」などの語ともつながりがあることから，画像分野の研究を主であることがわかる．また，なかでも「復元」との関連性が強く，画像の復元に関する研究がさかんであると推測できる.

本谷研： 「画像」とのつながりがあるため，画像分野の研究を主であることがわかる．また,「臓器」とのつながりから，医療分野の研究も行われていると推測できる.

舟橋研： つながりのある語のなかでも「VR」に強い関連性があり,「仮想」とのつながりもみられることから，VR 分野に関する研究を主であることがわかる.

梅崎研： 「計測」と「検査」とのつながりから，計測や検査を目的とした研究が多いとわかる．また,「認証」とのつながりから，セキュリティ分野の研究も行われていると推測できる．

白石研： 「通信」や「ネットワーク」などの語とのつながりから，通信分野の研究が行われていることがわかる．　また,「鍵」や「暗号」とのつながりから，通信のセキュリティ方面の研究がさかんであると推測できる．

岩田研： 「情報」「ネットワーク」とのつながりから，情報を扱った研究が行われているとわかる．また，データ数に対してつながりが少なく，関連性も弱いため，多種多様な研究を行っている，もしくはタイトルに特徴的な語を含まない分野の研究がさかんであると推測できる．

小田研： 「利他」との関連性が強く表れていることより，利他的行動に関する研究が行われていると推測できる．

平野研： 「信号」「処理」などとのつながりから，信号処理に関する研究が主であることがわかる．

松添研： 「分布」とのつながりから，分布を調査もしくは分布の分析などの研究が行われていると推測できる．

夏目研，クグレ研： これらの研究室はデータ数が極端に少ないため，関連性の強い語は見られなかった．

4.1.4　語の年度傾向

3.2.3 項で述べたように，語と各年度との関連性から，各年度ごとの研究テーマの傾向を分析する．語と卒業論文及び修士論文との関連性を表したネットワーク図を図 4.5 に示す．ただし，この図は程度の強い関係上位 100 個を表しており，また 10 回以上出現していない語はノードに含まれない．

まず,「音声」,「画像」などの出現回数の多い語はあらゆる年代において出現していることがわかる．一方，音声や画像と同じく出現回数が多かった「HMM」は 2009 年及び 2010 年，特に 2010 年において局所的に研究されていることが読み取れる．また,「カメラ」や「次元」という語は 2017 年にのみ関連性がみられることから，最新の研究でこれらの要素を使用するものが誕生したと推測できる．

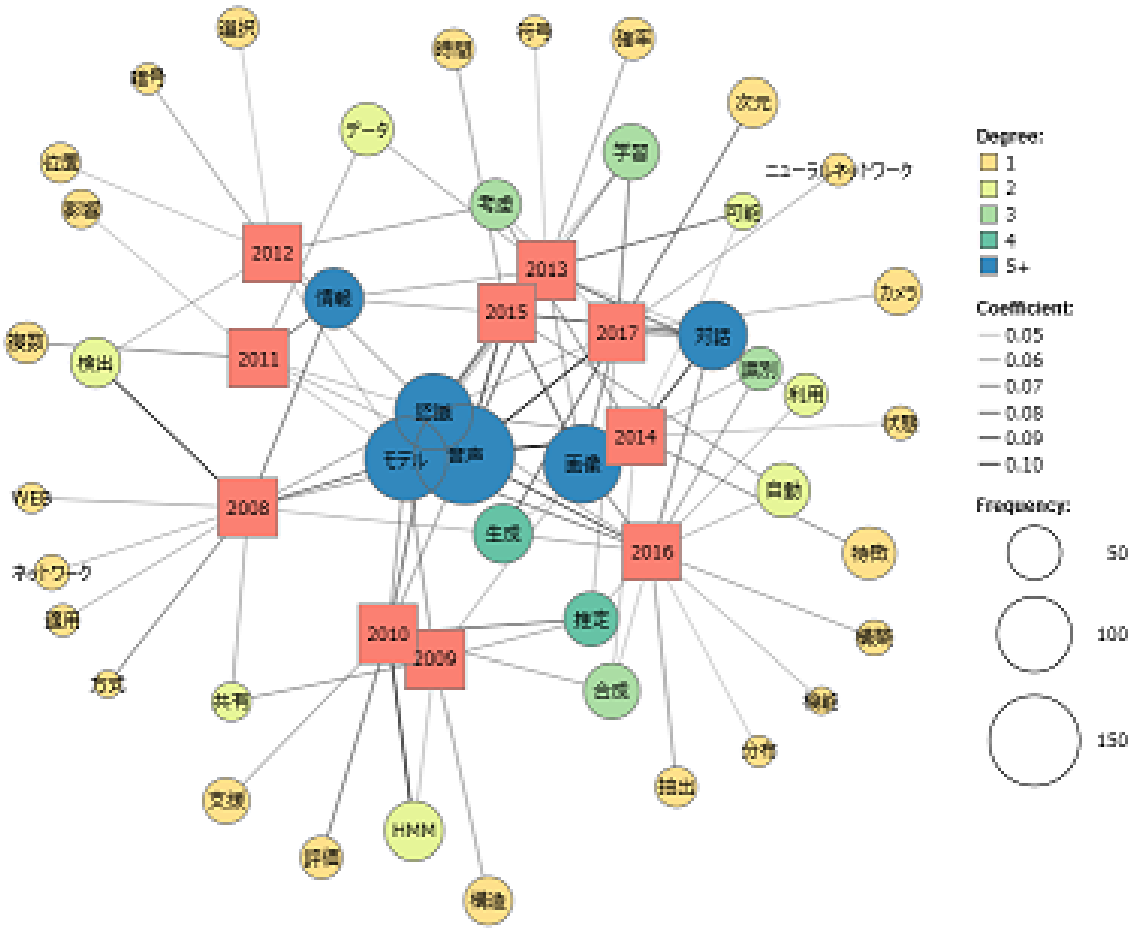

図 4.4: 語の年度傾向 (名工大)

4.1.5　卒業論文と修士論文の傾向

3.2.3 項で述べたように，卒業論文と修士論文のそれぞれの特徴を語との関連性から分析する．語と卒業論文及び修士論文との関連性を表したネットワーク図を図 4.5 に示す.「卒」は卒業論文を,「修」は修士論文を表す．ただし，この図は程度の強い関係上位 100 個を表しており，また 10 回以上出現していない語はノードに含まれない.

この図から目立った特徴はみられなかったが,「言語」「アルゴリズム」「ニューラルネットワーク」などの一般的に研究を行うのが困難であるものは修士研究で行われると言える.

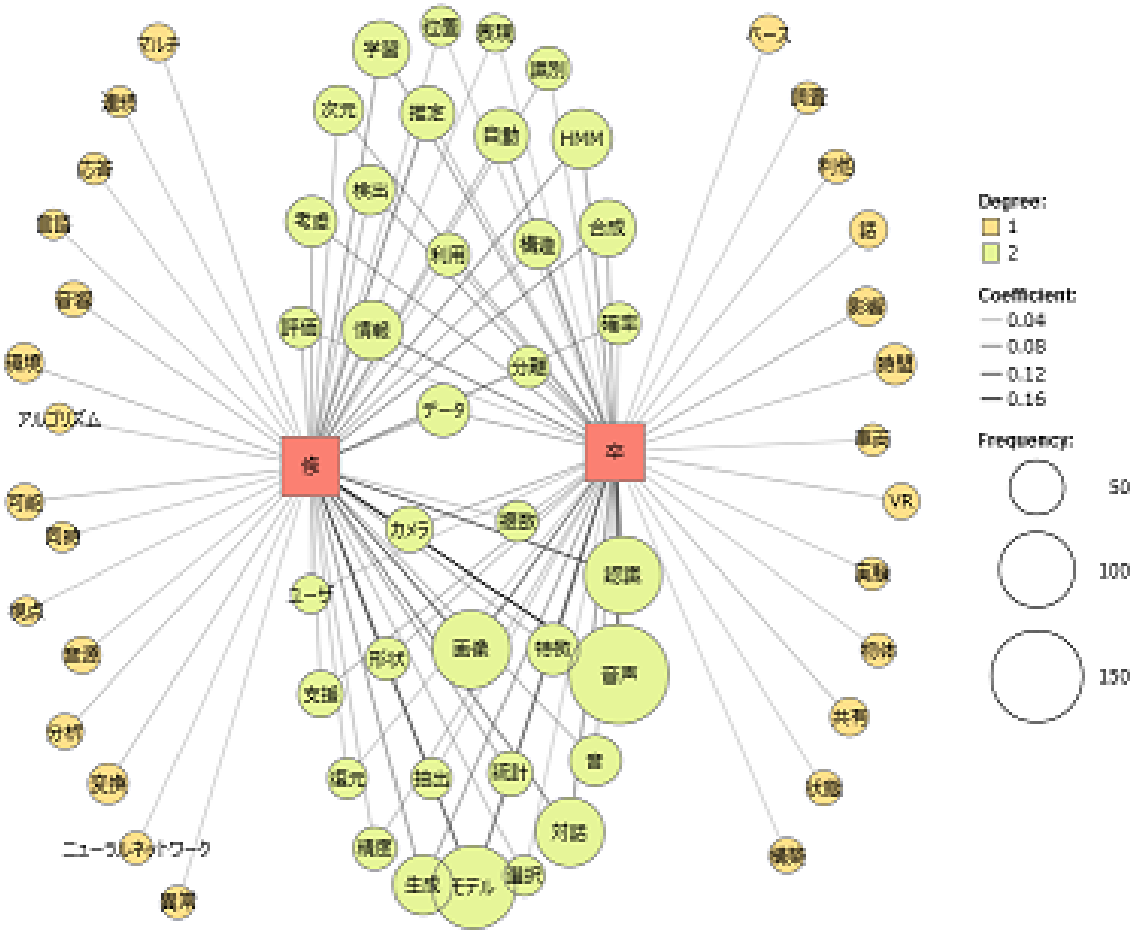

図 4.5: 卒業論文と修士論文 (名工大)

4.1.6　カテゴリ作成

3.2.4 項で述べたように，前項までの分析をもとに以下のカテゴリを作成した．

音声，画像，交通，医療，言語，VR，機械学習

さらに，音声分野のなかでも音楽要素に注目し,「音楽」というカテゴリを作成した．また，機械学習の詳しい手法について分析するため,「ニューラルネットワーク」「HMM」「ディープラーニング」の 3 つのカテゴリを作成した．各カテゴリが論文データの分類の指標とする語を表 4.3 に示す．なお，表の「ディープラーニング」欄の「畳み込む+ニューラルネットワーク」という部分は,「畳み込む」と「ニューラルネットワーク」という 2 つの語がひとつの論文タイトル内に同時に出現した場合を指す．

4.1.7　カテゴリごとの分析

3.2.4 項で述べたように，4.1.6 項で作成したカテゴリの設定に従い，各論文データをカテゴリに分類した．この項では，カテゴリに対して行った分析の結果と考察を述べる．

分類結果

各カテゴリに分類された論文データ数と全体に占める割合について，表 4.4 に示す．なお「分類なし」の項目は，どのカテゴリにも属さなかった論文データを指す．

この結果から，まず前項までの分析と同じように,「音声」カテゴリの全体に占める割合は極めて多い．また,「画像」カテゴリに関しても多くの割合を占めていることがわかる．次に,「音声」カテゴリと「音楽」カテゴリを比較すると,「音声」カテゴリに比べ「音楽」カテゴリが全体に占める割合は小さく,「音声」カテゴリのうちの約 2 割が「音楽」カテゴリである．また,「機械学習」カテゴリのうち,「HMM」カテゴリが約 5 割,「ニューラルネットワーク」カテゴリが約 4 割を占める．さらに,「ニューラルネットワーク」カテゴリのうち，約 4 割が「ディープラーニング」カテゴリに含まれている．

表 4.3: カテゴリに属する語

カテゴリ名	属する語
音声	音声，音響，音源，声質，発話，対話，聴覚，音素，音波，雑音，音量，音域， 音韻，補聴，無声，発生，発音，歌声，楽曲，音楽，楽譜，楽器，曲面，音程， 採譜，歌詞，演奏，音符，音律，歌唱，伴奏，編曲，合唱，作曲，基音，和声， 和音，打楽器，歌謡曲，三味線，音，譜，話しかける， スピーカ，マイク，オーディオ，ケプストラム，メロディ，バイオリン， アンサンブル, オーケストラ，ギター，クラシック，サビ，バンド， ピアノ，リズム，ビブラート，Sound
音楽	音楽，歌声，楽曲，楽譜，楽器，曲面，音程，採譜，歌詞，演奏，音符，音律， 歌唱，伴奏，編曲，合唱，作曲，基音，和声，和音，打楽器，歌謡曲，三味線， 音，譜，メロディ，バイオリン，アンサンブル，オーケストラ，ギター， クラシック，サビ，バンド，ピアノ，リズム，ビブラート
画像	画像，映像，動画，光源，視覚，撮像，視点，輝度，筆跡，外観，照度，陰影， 遠景，視野，写真，明暗，描画，撮影，写像，照明，解像度，絵， カメラ，シーン，ズーム，スクリーン，ディスプレイ，テクスチャ， ビデオモーションキャプチャ，HDR，HMD，Videography
交通	交通，車両，経路，路線，道路，渋滞，事故，車外，地図，車種，車載，案内， 移動，運転，配達，流通，駐車，道案内，車，路， カー，ドライバー，ルート，バス，GPS
医療	医療，臓器，治療，外科，皮質，医用，腫瘍，結節，静脈，病理，悪性，細胞， 動脈，脳波，肝臓，胸部，血管，障害，医学，衛生，患者，救急，血液，月経， 皮膚，病変，臨床，解剖，看護，介護，感染，療養，呼吸，癌，脳，肺，膵， 大脳皮質，インフルエンザ，カルテ，ポリープ，PET，PET/CT，FMD
言語	言語，語彙，辞書，言葉，単語，文字，漢字，索引，文章，文法，文脈，母語， 落語，翻訳，日本語，文字種，話し言葉，リンガル，Japanese
VR	仮想，複合現実，ヘッドマウントディスプレイ， バーチャル，グローブ，センサデータグローブ， VR，HMD，AR，MR
機械学習	機械学習，相関学習，強化学習，表現学習，決定木学習， マルチタスク学習，教師あり学習，教師なし学習，深層学習，自己組織化マップ， トランスダクション，トランスダクティブ推論，遺伝的プログラミング， マルコフ，クラスタリング，ベイジアンネットワーク，ニューラルネットワーク， ディープニューラルネットワーク，ニューラルネット， パルスニューラルネットワーク，リカレントニューラルネットワーク， ディープラーニング，GP，ELM，SVM，ILP，HMM，AI，Neural， CNN，DNN，CNNs，DNN-GMM，CombNET，CombNET-III
ニューラル ネットワーク	深層学習，自己組織化マップ，ニューラルネットワーク， ニューラルネット，ディープニューラルネットワーク， パルスニューラルネットワーク，ディープラーニング， リカレントニューラルネットワーク，Neural， CNN，DNN，CNNs，DNN-GMM，CombNET，CombNET-III
HMM	隠れマルコフ，HMM
ディープ ラーニング	深層学習，ディープラーニング，ディープニューラルネットワーク， リカレントニューラルネットワーク，畳み込む+ニューラルネットワーク， DNN，DNN-GMM，CNN，CNNs

表 4.4: カテゴリ分類結果 (名工大)

カテゴリ	論文データ数	全体に占める割合 (%)
音声	336	37.54
音楽	77	8.60
画像	190	21.23
交通	75	8.38
医療	88	9.83
言語	52	5.81
VR	47	5.25
機械学習	125	13.97
ニューラルネットワーク	53	5.92
HMM	63	7.04
ディープラーニング	22	2.46
分類なし	241	26.93

研究室ごとのカテゴリ分類

各研究室の論文データのうち，各カテゴリに分類されたものの数と研究室の論文全体に占める割合を表 4.5 及び表 4.6 に示す．ただし，データ数の少ない研究室は，特徴が過剰に出てしまうため，データ数が 10 以下であるクグレ，夏目，平野研は分析に含めない．また,「ニューラルネットワーク」「HMM」「ディープラーニング」は研究室のテーマの分析において冗長であるため，ここでは省略する．さらに，この表のデータをもとに作成した図を図 4.6 に示す．四角形の大きさが，表の割合に対応している．

この表及び図から以下のようなことが読み取れる．

- 音声分野を主として研究しているのは，黒柳，山本，徳田，北村，李研である．
- 音声分野を主として研究しているなかでも，北村研は特に音楽に関する研究を行っている．
- 音声分野を主として研究しているなかで，山本研はまったく音楽に関する研究を行っていない．
- 画像分野を主として研究しているのは，佐藤，本谷研である．

- 本谷研は医療分野も研究していることより，画像を用いた医療研究であると推測できる．
- 徳田研は機械学習を多く利用している．
- 山本研は交通カテゴリの研究も行っている．
- 岩田研はさまざまな分野を幅広く研究している．
- 言語について主に研究しているのは李研である．
- VR について研究しているのは主に舟橋研である．
- 小田，松添，白石，梅崎研は今回のカテゴリ分類で分類されなかった分野について主に研究していると推測できる．

表 4.5: 研究室ごとのカテゴリ分類結果 1(名工大)

研究室＼カテゴリ	音声	音楽	画像	交通
岩田	19 (19.79%)	1 (1.04%)	5 (5.21%)	16 (16.67%)
黒柳	47 (68.12%)	12 (17.39%)	3 (4.35%)	10 (14.49%)
佐藤	11 (9.73%)	11 (9.73%)	76 (67.26%)	13 (11.50%)
山本	24 (63.16%)	0 (0.00%)	3 (7.89%)	15 (39.47%)
舟橋	6 (8.57%)	2 (2.86%)	16 (22.86%)	6 (8.57%)
小田	0 (0.00%)	0 (0.00%)	6 (20.00%)	0 (0.00%)
松添	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
徳田	116 (73.42%)	19 (12.03%)	19 (12.03%)	1 (0.63%)
梅崎	1 (2.17%)	0 (0.00%)	9 (19.57%)	4 (8.70%)
白石	0 (0.00%)	0 (0.00%)	2 (5.26%)	3 (7.89%)
北村	40 (64.52%)	29 (46.77%)	4 (6.45%)	2 (3.23%)
本谷	10 (11.76%)	0 (0.00%)	44 (51.76%)	3 (3.53%)
李	60 (92.31%)	3 (4.62%)	0 (0.00%)	2 (3.08%)

表 4.6: 研究室ごとのカテゴリ分類結果 2(名工大)

カテゴリ / 研究室	医療	言語	VR	機械学習
岩田	20 (20.83%)	4 (4.17%)	0 (0.00%)	8 (8.33%)
黒柳	3 (4.35%)	2 (2.90%)	0 (0.00%)	8 (11.59%)
佐藤	6 (5.31%)	0 (0.00%)	4 (3.54%)	4 (3.54%)
山本	0 (0.00%)	1 (2.63%)	0 (0.00%)	0 (0.00%)
舟橋	4 (5.71%)	0 (0.00%)	42 (60.00%)	5 (7.14%)
小田	1 (3.33%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
松添	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (9.09%)
徳田	0 (0.00%)	15 (9.49%)	0 (0.00%)	69 (43.67%)
梅崎	3 (6.52%)	3 (6.52%)	0 (0.00%)	5 (10.87%)
白石	1 (2.63%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
北村	3 (4.84%)	10 (16.13%)	0 (0.00%)	11 (17.74%)
本谷	47 (55.29%)	1 (1.18%)	1 (1.18%)	4 (4.71%)
李	0 (0.00%)	16 (24.62%)	0 (0.00%)	8 (12.31%)

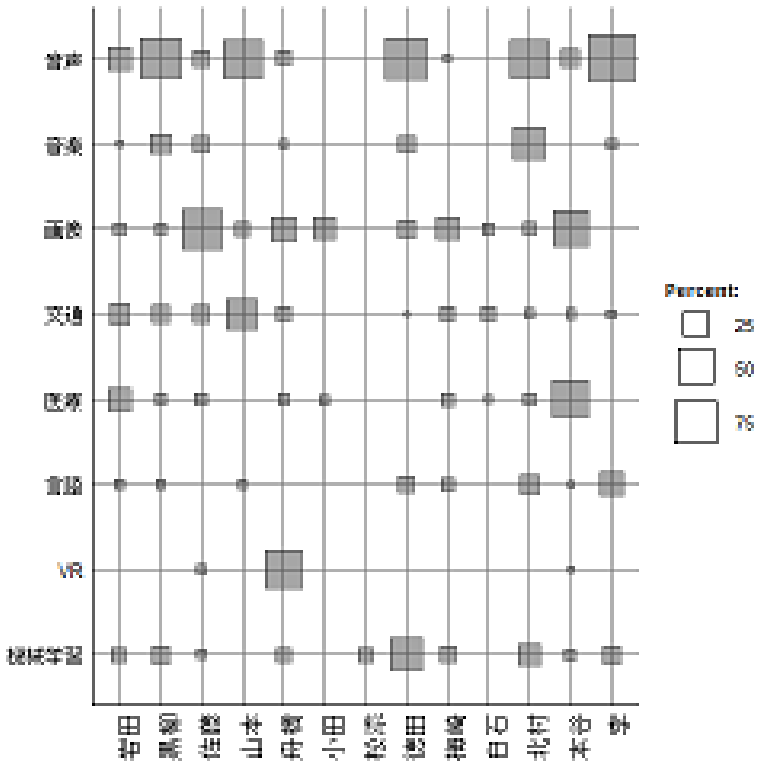

図 4.6: 研究室ごとのカテゴリ特徴 (名工大)

機械学習の手法の年度推移

「機械学習」「ニューラルネットワーク」「HMM」「ディープラーニング」のカテゴリの年度ごとの研究全体に占める割合を見ることで，大学研究における機械学習の手法の推移について探る．これを折れ線グラフで表したものを図 4.7 に示す．

この図から，まず 4.1.5 項でも述べたように，2010 年に HMM が急激に流行していたことがわかる．また，2012 年以降ニューラルネットワークを用いた研究は減少している．2015 年以降はディープラーニングの台頭によりニューラルネットワークを用いた研究も再び増加し，一方で HMM を用いた研究はディープラーニングの登場により減少している．

ディープラーニング研究における大きな出来事として，2012 年に，ディープラーニングの概念の提唱者である Hinton らが画像認識コンテストで圧勝した，というものがある．図より，名工大においてディープラーニング研究が行われ始めたのが 2015 年からであるので，ディープラーニングが注目されてから学生が研究に取り入れるまで 3 年かかったということがわかる．これは，ディープラーニングという手法を画像や音声の研究に応用できるようになるまで時間がかかったのだと考えられる．

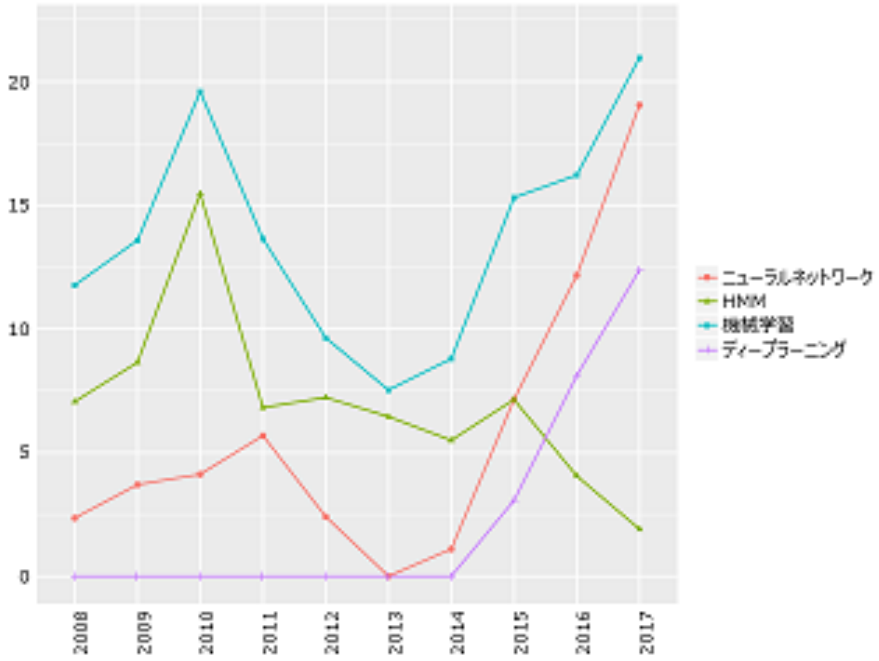

図 4.7: 機械学習手法の年度推移

VR 研究の年度推移

VR 研究の推移についても探る.「VR」カテゴリの年度ごとの研究全体に占める割合を折れ線グラフで表したものを図 4.8 に示す.

VR に関する研究は，小さな増加と減少を繰り返していることが読み取れる．VR 研究における大きな出来事としては，2013 年に「Oculus Rift」の開発キットが公開されている．図の 2013 年に注目すると，前年より少し増加していることがわかる．また，2016 年は一般向け VR が普及し,「VR 元年」と呼ばれた年である．図の 2016 年では前年より減少しているが，2017 年には増加していることより，これから VR に関する研究が増加することが期待できる.

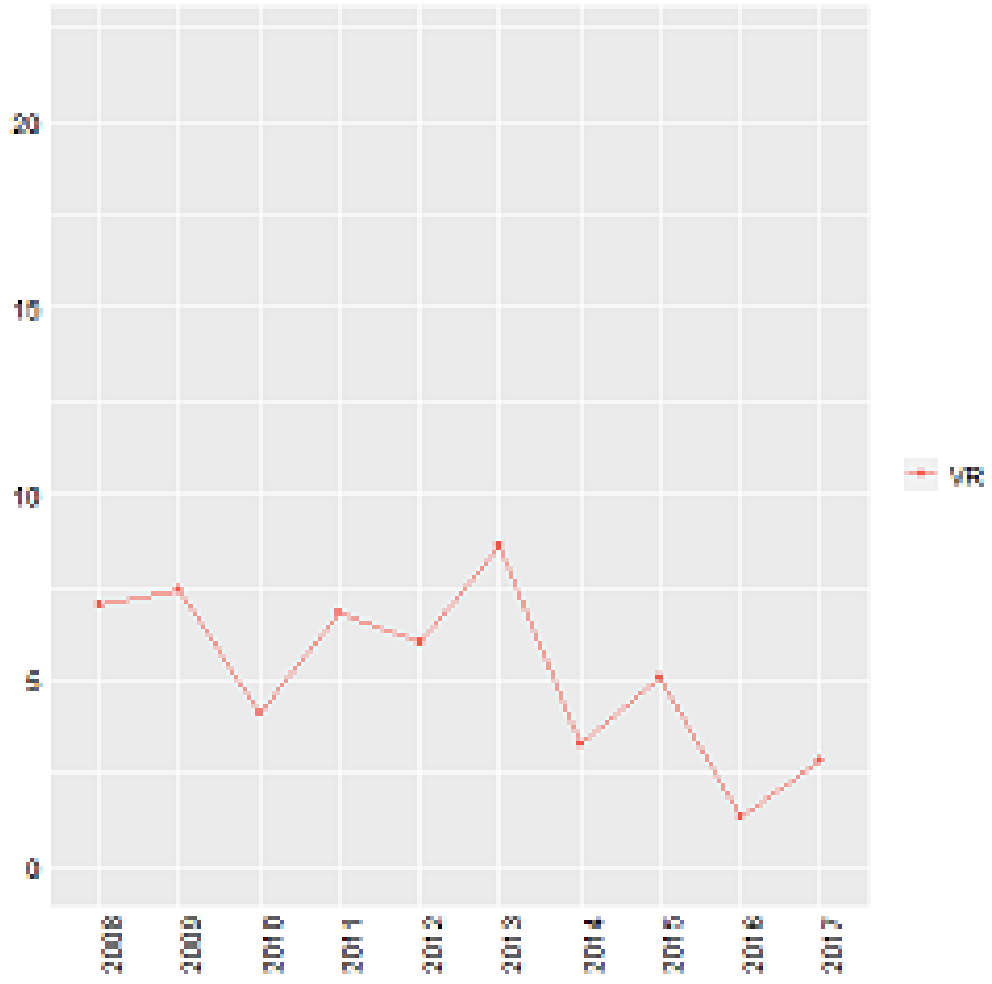

図 4.8: VR 研究の年度推移

4.2　東京大学のデータを用いた分析

本節では，東京大学のデータに対して行った分析の結果と，それに対する考察を述べる．また，4.1 節の結果と比較する．

4.2.1　抽出された語

4.1.2 項と同じように語の取捨選択を行った後，語の抽出を行った．表 4.7 に，抽出された語の中で出現回数が多いもの上位 20 個を示す．また, この抽出結果をもとに共起ネットワークを作成した．この共起ネットワークを図 4.9 に示す．

表 4.7: 語の抽出結果 (東大)

抽出語	出現回数
行動	58
ロボット	51
ヒューマノイド	42
運動	39
計測	34
操作	31
生成	30
認識	30
モデル	29
支援	29
環境	27
画像	26
制御	26
動作	26
物体	26
センサ	25
情報	25
推定	25
神経	21
学習	20

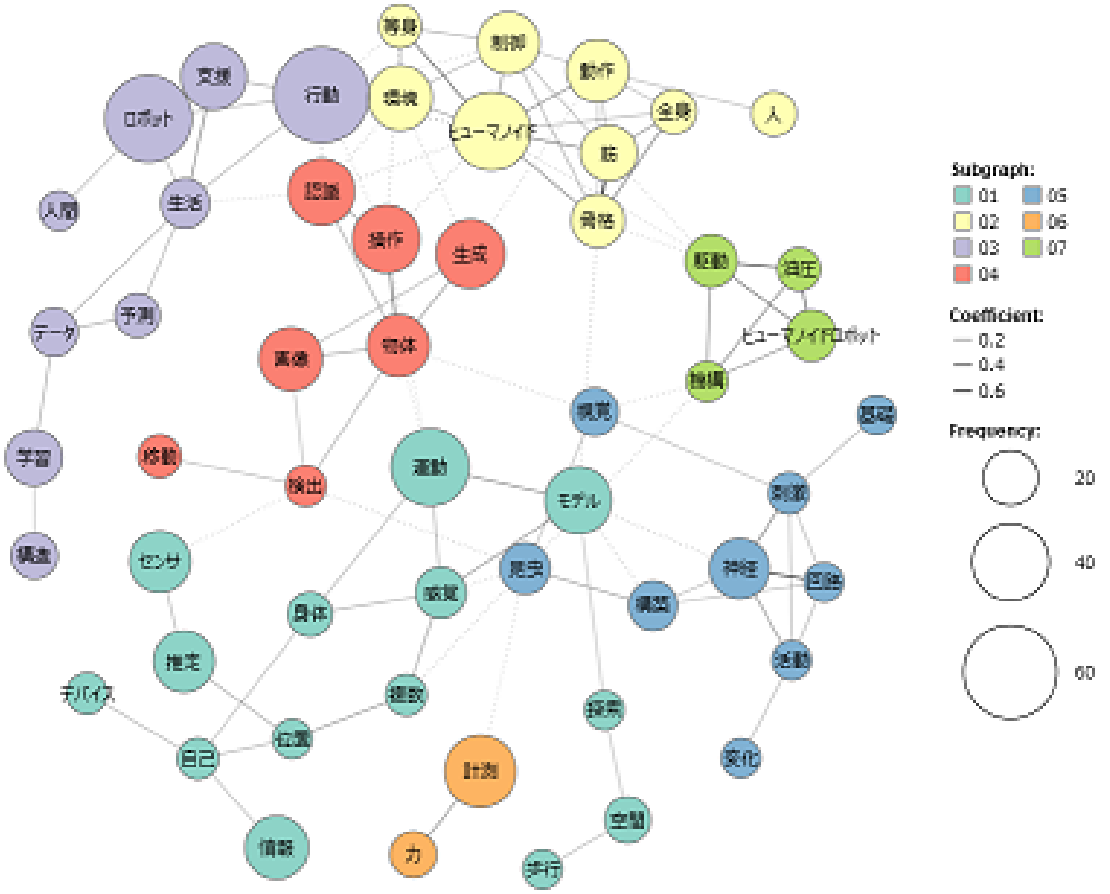

図 4.9: 共起ネットワーク (東大)

語の抽出結果より,「ロボット」「ヒューマノイド」の出現回数が多いことから，ロボット分野の研究がさかんであるとわかる．このことを踏まえると,「行動」「運動」「操作」などのロボットに関する語が多く出現していることがわかる．

また，共起ネットワークより,「ロボット」に注目すると,「生活」「支援」とのつながりから，ロボットによる生活支援の研究が行われていると推測できる．また,「ヒューマノイド」に注目すると,「筋」「骨格」「等身」などのロボットの構成に関する要素や,「制御」「動作」などのロボットの動きに関する要素がみられる．

4.1.2 項の結果と比較すると，名工大のデータでは音声分野や画像分野が目立っていたのに対し，このデータではロボット分野の研究が目立つ．また，音声分野に関する語はあまりみられなかった．ただし,「画像」という語は多く抽出されており，共通している部分も存在するということがわかる．

4.2.2　語と研究室の関係

4.1.3 項と同様に，語と東大の各研究室との関連性について分析を行った．語と研究室との関連性を表したネットワーク図を図 4.11 に示す．ただし，この図は程度の強い関係上位 70 個を表しており，また 10 回以上出現していない語はノードに含まれない．

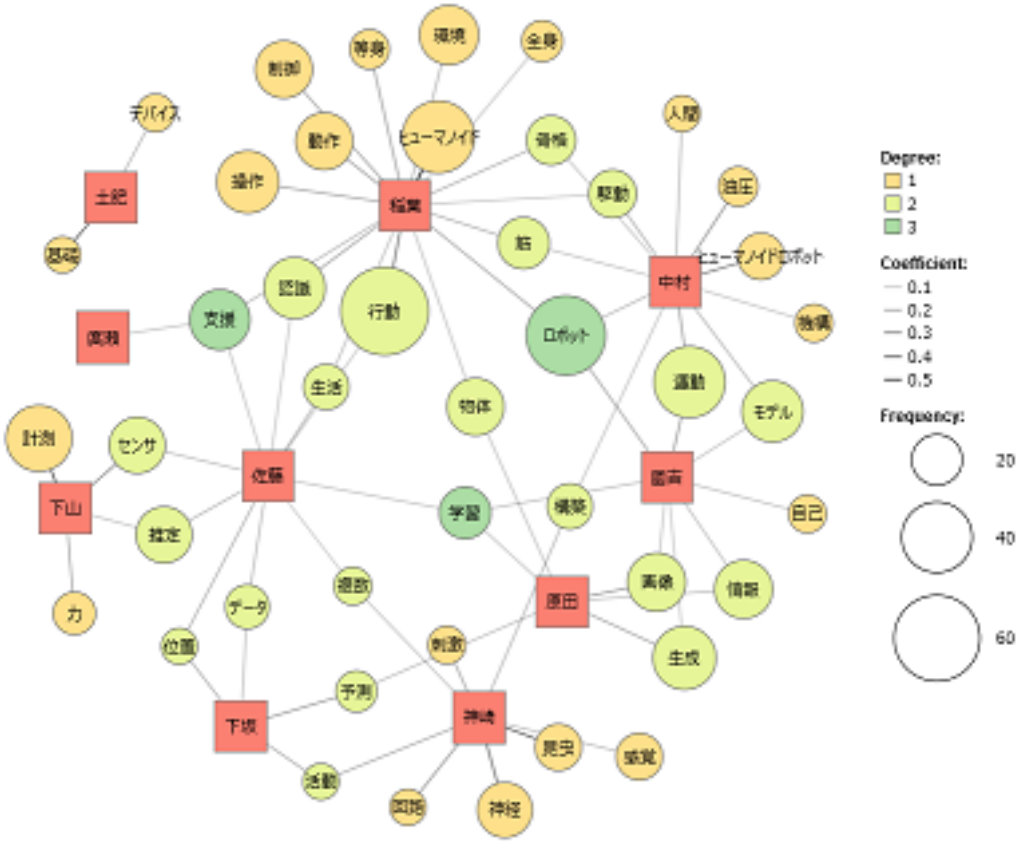

図 4.10: 研究室傾向 (東大)

この図から，各研究室について以下のようなことが読み取れる．

稲葉研：「ロボット」「ヒューマノイド」とのつながりがあり,「動作」「筋」などのロボットに関連する語とのつながりも多くみられるため，ロボット分野を主に研究していることがわかる．特に「ヒューマノイド」との関連性が強くでているため，ロボットのなかでもヒューマノイドを多く扱っていると推測できる．

下山研：「力」「計測」とのつながりから，力の計測などの計測分野を主に研究し

ているとわかる．また,「センサ」ともつながりがあることから，センサを用いた計測が行われていると推測できる．

原田研： 「画像」とのつながりから，画像分野の研究を行っていることがわかる．また,「学習」ともつながりがあることから，機械学習を用いて画像の研究を行っていると推測できる．

佐藤研： 「データ」「学習」とのつながりから，機械学習を用いた研究を行っていると考えられる．また,「生活」「行動」「推定」とのつながりより，機械学習を用いて生活行動を推定していると推測できる．

神崎研： 「神経」「回路」とのつながりから，神経回路に対する研究が行われていることが推測できる．また,「昆虫」との関連性が強く，昆虫に関する研究がさかんであるといえる．

中村研： 「ロボット」「ヒューマノイドロボット」とのつながりがあることから，ロボット分野を主に研究していることがわかる．なかでも,「駆動」「骨格」「油圧」「機構」など，ロボットの構成に関わる研究が多く行われていることが推測できる．

土肥研： 「基礎」との関連性が強いことから，基礎部分に関しての研究が行われているとわかる．

國吉研： 「運動」「モデル」「構築」とのつながりから，運動モデルに関する研究が行われていることが推測できる．また,「画像」ともつながりがあるため，画像分野の研究も行っていることがわかる．

廣瀬研： データ数に対してつながりが少ないことより，幅広い分野で研究を行っていると考えられる．また,「支援」とのつながりから，なかでもユーザを支援する研究が多いと推測できる．

下坂研： 「活動」「位置」「予測」とのつながりから，活動や位置などを予測する研究が行われているとわかる．

正宗研： データ数が極端に少ないため，関連性の強い語は見られなかった．

4.2.3　カテゴリ作成

名工大の結果と比較するために，4.1.6 項で作成したカテゴリを使用する．使用するカテゴリは以下の 5 つである．

画像，音声，VR，医療，機械学習

なお，他のカテゴリに関しては，東京大学のデータの分析に対して冗長であるため除く．また，4.2.1 項で行った分析結果より，新たに「ロボット」というカテゴリを作成した.「ロボット」カテゴリが分類の指標とする語を表 4.8 に示す.

表 4.8: ロボットカテゴリに属する語

カテゴリ	属する語
ロボット	ロボット，ヒューマノイド，ヒューマノイドロボット，モジュラーロボット，キメラロボット，モビリティロボット，マニピュレーション，パーソナルモビリティロボット

4.2.4　カテゴリごとの分析

4.2.3 項で作成したカテゴリの設定に従い，各論文データをカテゴリに分類した．この項では，カテゴリに対して行った分析の結果と考察を述べる．

分類結果

各カテゴリに分類された論文データ数と全体に占める割合について，表 4.9 に示す．なお「分類なし」の項目は，どのカテゴリにも属さなかった論文データを指す.

この結果から，まず前項までの分析どうり,「ロボット」カテゴリが大きな割合を占めていることがわかる．また,「画像」カテゴリも多くの研究が行われていることがわかる．一方，残りのカテゴリは占める割合が小さい.

表 4.4 の結果と比較すると，このデータで「ロボット」が全体に占める割合に対して，名工大のデータで「音声」が全体に占める割合は大きく,「画像」が占める割合は小さい．このことから，東大のデータにおけるロボット研究は，名工大のデータにおける画像分野よりもさかんに研究されているが，名工大のデータにおける音声

表 4.9: カテゴリ分類結果 (東大)

カテゴリ	論文データ数	全体に占める割合 (%)
ロボット	113	26.40
画像	65	15.19
音声	27	6.31
VR	20	4.67
医療	30	7.01
機械学習	14	3.27
分類なし	189	44.16

分野よりはさかんではないということがわかる．また，東大のデータでは分類されなかった論文が多いため，このデータを詳しく分析するためには新たなカテゴリを作成することが課題である．

研究室ごとのカテゴリ分類

各研究室の論文データのうち，各カテゴリに分類されたものの数と研究室の論文全体に占める割合を表 4.10 及び 4.11 に示す．ただし，データ数の少ない研究室は特徴が過剰に出てしまうため，データ数が 10 以下である下坂，正宗研は分析に含めない．さらに，この表のデータをもとに作成した図を図 4.6 に示す．

表 4.10: 研究室ごとのカテゴリ分類結果 1(東大)

研究室＼カテゴリ	ロボット	画像	音声
稲葉	59 (90.77%)	4 (6.15%)	2 (3.08%)
下坂	0 (0.00%)	0 (0.00%)	1 (16.67%)
下山	1 (2.04%)	1 (2.04%)	5 (10.20%)
原田	0 (0.00%)	13 (44.83%)	2 (6.90%)
佐藤	3 (10.00%)	6 (20.00%)	0 (0.00%)
神崎	2 (5.56%)	3 (8.33%)	3 (8.33%)
正宗	1 (25.00%)	0 (0.00%)	0 (0.00%)
中村	27 (49.09%)	7 (12.73%)	1 (1.82%)
土肥	1 (5.88%)	6 (35.29%)	3 (17.65%)
廣瀬	0 (0.00%)	14 (20.29%)	6 (8.70%)
國吉	19 (27.94%)	11 (16.18%)	4 (5.88%)

表 4.11: 研究室ごとのカテゴリ分類結果 2(東大)

研究室＼カテゴリ	VR	医療	機械学習
稲葉	0 (0.00%)	0 (0.00%)	0 (0.00%)
下坂	0 (0.00%)	0 (0.00%)	0 (0.00%)
下山	0 (0.00%)	0 (0.00%)	0 (0.00%)
原田	0 (0.00%)	2 (6.90%)	4 (13.79%)
佐藤	0 (0.00%)	0 (0.00%)	3 (10.00%)
神崎	1 (2.78%)	12 (33.33%)	0 (0.00%)
正宗	0 (0.00%)	2 (50.00%)	0 (0.00%)
中村	0 (0.00%)	2 (3.64%)	1 (1.82%)
土肥	0 (0.00%)	6 (35.29%)	0 (0.00%)
廣瀬	16 (23.19%)	0 (0.00%)	0 (0.00%)
國吉	3 (4.41%)	6 (8.82%)	6 (8.82%)

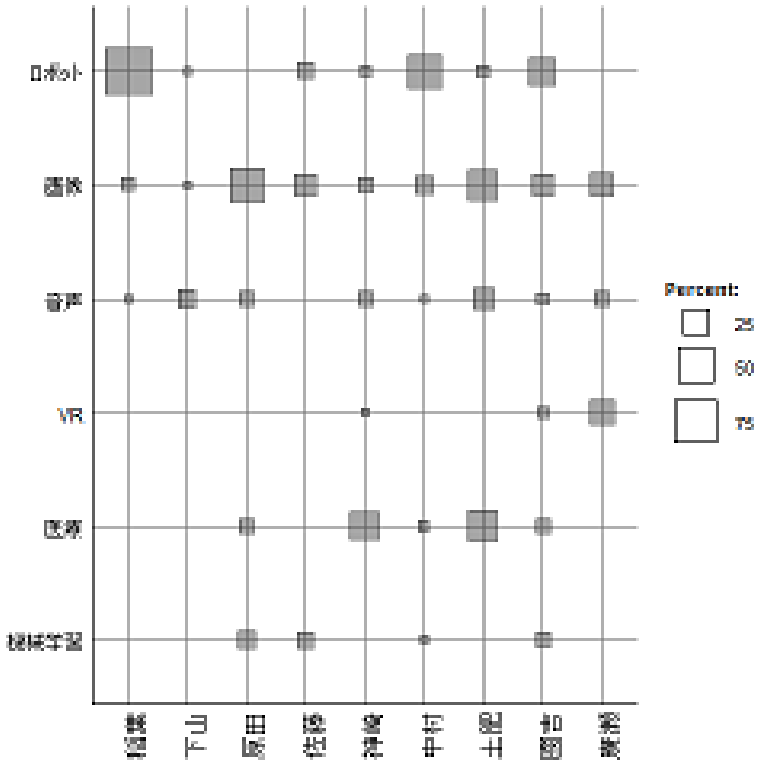

図 4.11: 研究室ごとのカテゴリ特徴 (東大)

この表及び図から以下のようなことが読み取れる．

- ロボット分野を主に研究しているのは，稲葉，中村，國吉研である．
- ロボット分野を研究しているなかでも，稲葉研はロボット分野に大きく力をいれている．
- 原田研は画像分野を主に研究している．
- 医療関係の研究に力をいれているのは，神崎研と土肥研である．
- 土肥研では，画像分野の研究も行っていることから，画像を用いた医療研究を行っていると推測できる．
- 廣瀬研では，VR 関係の研究を行っている．
- 國吉研では，ロボット分野の他にも様々な分野の研究を行っている．
- 機械学習を用いた研究は，主に原田研，佐藤研，國吉研で行われている．
- 佐藤，下山研は，今回のカテゴリ分類で分類されなかった分野の研究を中心におこなっていると推測できる．

4.2.5　分析の正当性の確認

分析結果の正当性を確かめるため，各研究室の Web ページを参考にする．例として，図 4.12 に稲葉研究室の研究紹介ページ [9] を示す．この Web ページから，稲葉研究室はロボットの研究を中心に行っていることがわかる．これは，分析結果と一致する．このように，各研究室ごとに Web ページ [10][11][12][13][14][15][16] で分析の結果と比較したところ，研究テーマが分析結果と大きく違うことはなかった．ただし，佐藤，下坂，正宗研究室に関しては Web ページが存在しなかったため，比較を行っていない．

図 4.12: 稲葉研究室 HP

第5章　むすび

本研究では，名古屋工業大学及び東京大学の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，大学で行われている研究のテーマの可視化と分析を行った．第2章では，本研究で用いたテキストマイニングの概要及び全体の流れと，文の語への分解法，出力する図である共起ネットワークについて説明した．第3章では，用いた研究論文のタイトルデータの概要と，行う分析の手法について述べた．第4章では，実際に分析を行い，その結果について分析・考察を行った．

その結果，各研究室ごとの研究テーマの傾向を捉えることができる図を作成できた．なかでも，カテゴリ分類を用いた研究室ごとの特徴を分析した図より，各研究室のおおまかな特徴を捉えることができた．さらに，研究室と語の関係を表したネットワーク図を利用することで，各研究室の具体的な研究テーマについても知ることができた．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．

今後の課題としては，用いるデータの量や種類を増やし，大学ごとの研究テーマの傾向をより詳しく分析したり，同じ大学の学科ごとの特徴を分析することなどが挙げられる．また，第4章で作成したカテゴリを，Web上のデータなどを参考に関連する語を抽出し，自動的に作成することができれば，さらに分析の汎用性が増すだろう．他にも，年度推移の分析から大学における今後の研究テーマのトレンドがわかれば，研究方針の参考になるだろう．そして将来的には，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，受験生が短時間で情報を得ることにより，進路を決める指針となることを期待している．

謝辞

本研究を進めるにあたって，日頃から多大な御尽力を頂き，御指導を賜りました名古屋工業大学，舟橋健司 准教授，伊藤宏隆 助教に心から感謝致します．最後に，本研究に多大なご協力頂きました舟橋研究室諸氏に深く感謝致します．

参考文献

[1] 原圭司, 高橋健一, 上田祐彰:“ ベイジアンネットワークを用いた授業アンケートからの学生行動モデルの構築と考察 ”, 情報処理学会論文誌, Vol.51,No.4,pp.1215-1226, 2010.

[2] 伊藤圭佑, 舟橋健司, 伊藤宏隆:“ データマイニングによる要注意学生の発見に関する研究 ”, 平成２５年度名古屋工業大学卒業論文, 2013.

[3] 渡会純一:“ ピアノ演奏指導に関する考察と展望 -「表現技術Ⅰ(音楽)」履修学生への意識調査から- ”教職研究, Vol.2017,pp.83-101, 2018.

[4] 原以起, 奥野圭太郎, 奥野卓司:“ テキストマイニングによる「未来社会」解読の試行 ”, 関西学院大学先端社会研究所紀要, vol.15,pp.91-105, 2018.

[5] 樋口耕一:“ テキスト型データの計量的分析：2つのアプローチの峻別と統合 ”, Sociological theory and methods, vol.19,No.1,pp101-115, 2004.

[6] 日経リサーチ: “テキストマイニング”, https://www.nikkei-r.co.jp/glossary/id=1602/, 2019-1-23 参照.

[7] 東京大学機械情報工学科: “卒業論文題目”, http://www.kikaib.t.u-tokyo.ac.jp/komaba/thesis/, 2019-1-22 参照.

[8] 樋口耕一:“ KH Coder:計量テキスト分析・テキストマイニングのためのフリーソフトウェア ”, http://khcoder.net/, 2019-1-22 参照.

[9] 稲葉雅幸:“ JSK -Research- ”, http://www.jsk.t.u-tokyo.ac.jp/research-j.html, 2019-2-1 参照.

[10] 土肥健純:“ 東京大学 大学院 情報理工学系研究科 ”, http://www.i.u-tokyo.ac.jp/edu/course/m-i/showcase/show_3.shtml, 2019-2-1 参照.

[11] 神崎亮平：“ Kanzaki Lab ”，http://www.brain.rcast.u-tokyo.ac.jp/，2019-2-1 参照.

[12] 中村仁彦：“ Research -Nakamura&Yamamoto Lab ”，http://www.ynl.t.u-tokyo.ac.jp/wp/research/，2019-2-1 参照.

[13] 廣瀬通孝：“ プロジェクト—廣瀬・谷川・鳴海研究室 ”，http://www.cyber.t.u-tokyo.ac.jp/ja/projects/，2019-2-1 参照.

[14] 國吉康夫：“ ISI Laboratory –The University of Tokyo ”，http://www.isi.imi.i.u-tokyo.ac.jp/?lang=ja，2019-2-1 参照.

[15] 下山勲：“ Shimoyama Lab, The University of Tokyo,Japan ”，http://www.leopard.t.u-tokyo.ac.jp/index.html，2019-2-1 参照.

[16] 原田達也：“ https://www.mi.t.u-tokyo.ac.jp/members/ ”，https://www.mi.t.u-tokyo.ac.jp/research/，2019-2-1 参照.

付録A KH Coderについて

本研究では，テキストマイニングによる分析を行うためにフリーソフトウェアである KH Coder を利用している．KH Coder は読み込んだテキスト型データに対して，データ中から語を抽出し，多く出現していた語の確認や，語の特徴の図示を行うことができるソフトである．本研究で用いた KH Coder は ver3.0.0.0 である．また，KH Coder による分析に適したデータを作成するため，Microsoft 社の Excel2010 を利用している．

実際の手順

4.1 節で行った分析の手法を例に使い方を紹介していく．

1. あらかじめ Excel で「論文タイトル」「年度」「研究室」「卒業論文か修士論文か(卒/修)」を対応させたデータセットを作成する．なお，Excel データの 1 行目には各要素の名前を入力する必要がある．
2. KH Coder を起動し，メニューの「プロジェクト → 新規」から，データを読み込ませる．このとき,「分析対象とする列」を「論文タイトル」に設定する．
3. メニューの「前処理 → 分析対象ファイルのチェック」で，KH Coder の規定にそぐわない部分を変更する．
4. 「前処理 → 語の取捨選択」から，3.2.2 項で解説した語の取捨選択を設定する．
5. 「前処理 → 前処理の実行」で，語の抽出を行う．
6. 「ツール → 抽出語 → 抽出語リスト」から，抽出された語の出現回数を確認する．

7. 「ツール → 抽出語 → 共起ネットワーク」を設置し，共起ネットワークのオプション画面を開く．初期設定から「最小文書数」を 10,「描画する共起関係の選択」を上位 100 に変更し,「強い共起関係ほど濃い線に」にチェックをいれ,「半透明の色」のチェックを外し，実行する．これにより，図 4.2 が出力される．

8. 7 と同様にオプションを選択した後,「共起関係の種類」を「語-外部変数・見出し」に変更し,「外部変数・見出し」を「研究室」に変更することで図 4.3 が出力される．同様に,「外部変数・見出し」を「年度」に変更することで図 4.4,「卒/修」に変更することで図 4.5 が出力される．

9. あらかじめ，表 4.3 の内容を，コーディングルールとして txt ファイルで作成しておき,「ツール → コーディング → 単純集計」から，各論文をカテゴリに分類し，表 4.4 の結果が得られる．

10. 「ツール → コーディング → クロス集計」から,「クロス集計」を「研究室」に設定し集計することで，表 4.5 の結果が得られ,「マップ → バブル」を設置することで，この結果を図示することができる．また,「クロス集計」を「年度」に変更することでカテゴリの年度傾向をみることが可能である．

以上が本研究で行った分析の一例についての説明である．コーディングルールの設定方法は KH Coder 付属のマニュアルを参考されたい．