

指導教員（主査）：山本祐輔 講師

副査：福田直樹 准教授

2018 年度 静岡大学情報学部 卒業論文

笑えるウェブ情報検索のためのクエリ推薦

静岡大学 情報学部 CS プログラム 所属

学籍番号 70510016

梅田 浩郎

2019 年 2 月 14 日

概要

本稿では，ウェブ検索プロセスにおいて笑いの機会をつくる一つの方法として，笑いを誘発するクエリの推薦手法を提案する．提案手法は，クエリと組み合わせると笑いを誘発する語をクエリ入力時に推定し推薦する．提案手法は，クエリに対して意外な組み合わせであり，クエリの極性と異なる極性を持つ単語を，クエリと組み合わせる単語とする．既存のクエリ推薦手法の目的は，入力クエリと関連性が高いクエリを推薦することによって，検索行動の効率化を図ることである．提案手法を用いることで，ウェブ検索中のユーザーに，効率性に着目した情報検索システムでは得がたい笑いの機会を提供することが可能となる．

目次

第 1 章	はじめに	5
第 2 章	関連研究	7
2.1	笑いのメカニズム	7
2.2	笑いを支援するシステム	7
2.3	クエリ推薦	8
第 3 章	笑いを誘発するクエリの要件	9
第 4 章	提案内容	11
4.1	関連語の収集	11
4.2	意外性	12
4.3	対立性	15
4.4	笑える語の組み合わせのランキング	15
第 5 章	評価実験	16
5.1	比較手法	16
5.2	評価尺度	16
5.3	評価用クエリ	17
5.4	評価者	17
5.5	実験手順	18
第 6 章	結果	19
第 7 章	考察	26
第 8 章	おわりに	28

目次

1.1	既存のクエリ推薦システム (左) と提案システム (右)	6
4.1	同位語度の計算イメージ	13
4.2	連想度の計算イメージ	14
6.1	アンパンマンの適合率の推移	22
6.2	ミッキーマウスの適合率の推移	22
6.3	きかんしゃトーマスの適合率の推移	22
6.4	ムーミンの適合率の推移	22
6.5	リラックマの適合率の推移	23
6.6	野球の適合率の推移	23
6.7	サッカーの適合率の推移	23
6.8	ボクシングの適合率の推移	23
6.9	テニスの適合率の推移	24
6.10	ゴルフの適合率の推移	24
6.11	カレーライスの適合率の推移	24
6.12	ラーメンの適合率の推移	24
6.13	ハンバーグの適合率の推移	25
6.14	肉じゃがの適合率の推移	25
6.15	炒飯の適合率の推移	25

表目次

5.1	評価用クエリ一覧	17
6.1	$P@5$, $P@10$, $P@15$, AP , MAP の評価結果	20
6.2	提案手法の出力結果の上位 5 件 (○は笑えると判定された組み合わせ, × は笑えないと判定された組み合わせ)	21

第 1 章

はじめに

人は笑いを求めており，多くの場面で笑うための情報を収集している．情報を集めるためのメディアとして，インターネットが利用されている．特に趣味・娯楽に関する情報については，テレビなどの他メディアよりインターネットのほうが利用されることが多い [1]．インターネットの利用目的として，メールやソーシャルメディアの利用，ウェブサイトの閲覧，動画共有サイトでの動画視聴などがあげられる．特に，ソーシャルメディアや動画共有サイトなど，利用者の意見や感想を他人と共有できるインターネットメディアの利用時間は長い．ソーシャルメディアに投稿された意見や感想には，笑いを表現するさまざまな用語が用いられている．例えば，“笑”や“w”といった用語があげられる．“w”は“笑い”を省略して入力するために生まれた．笑いに関する新たな文字表現が生まれるほど，人は笑いを求めている．また，笑いはわれわれ人間のこころや身体に良い影響を与えることが研究によって明らかになっている．例えば，笑いにはストレスや痛みの軽減，免疫力を向上させるといった効果があることが知られている [2]．それゆえ，日常生活において笑いの機会を増やすことは重要である．

ウェブ検索を用いて笑える情報を取得することは予想以上に難しい．例えば，“アンパンマン”について笑えるウェブページを検索したい場合，単純な方法として，“アンパンマン 笑い”というクエリを検索エンジンに入力することが考えられる．しかし，既存のウェブ検索エンジンは入力されたキーワードを含むウェブページを検索結果として返すため，「笑い」という言葉を含んでいるページが必ずしも笑える内容を含んでいるとは限らない．そのため，笑える情報をウェブから検索するためには，笑える情報を導くクエリを考えて検索エンジンに入力する必要がある．しかし，検索ユーザにとってはどのようなトピックが笑える情報を含んでいるかは未知であるため，笑える情報を導くクエリを作成することは難しい．また，仮に笑えるネタを思いつき，それをクエリとしてウェブ検索したとして

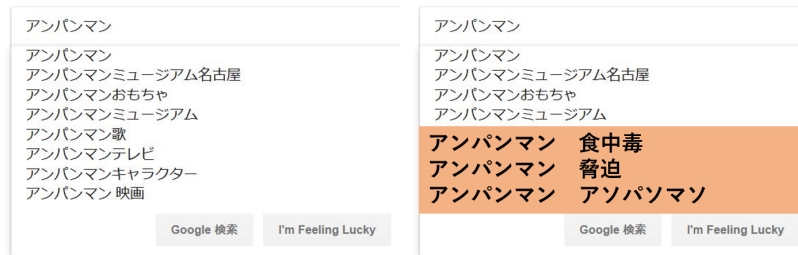


図 1.1 既存のクエリ推薦システム (左) と提案システム (右)

も、そのネタに関連するコンテンツがウェブに存在しないことも考えられる。

本稿では、ウェブ検索エンジンのクエリ入力時に、笑いを誘発するクエリを推薦するシステムを提案する。提案手法は、クエリと組み合わせると意外性を感じる単語で、かつクエリの極性と異なる極性を持つ単語を、笑いを誘発するクエリとして推薦する。提案手法は、クエリ入力時に笑いを誘発するクエリを提示することで、検索プロセス中に検索結果に依存せずに笑いの機会を提供することを目的とする。提案手法はウェブから笑える情報を探すことを最終目的としない。

図 1.1 に提案システムのイメージを示す。提案システムは、既存のウェブ検索エンジンが提供しているクエリ推薦機能のように、検索ボックスの下に入力されたクエリに関連し、かつ笑いを誘発するクエリを推薦する。図 1.1 では、“アンパンマン”を検索クエリとした場合である。笑いを誘発するクエリとして“アンパンマン 食中毒”といったクエリを推薦している。

本稿の以降の構成は以下の通りである。2 章では関連研究について述べる。3 章では笑いの条件について述べる。4 章では笑いを誘発するクエリの推薦方法について述べる。5 章では実験内容について述べ、6 章で実験結果を示し、7 章で考察を述べる。最後に 8 章でまとめを述べる。

第 2 章

関連研究

2.1 笑いのメカニズム

笑いを発生させる条件についての研究はいくつか行われている。マグロウらの笑いの理論「無害と逸脱」では、人が不快と感じる物事(逸脱)を不快と感じないようにする(無害化)と、笑いが発生するとしている [3]。マグロウの理論は、提案手法の笑いの条件を考察する上で参考にした。嶋田は複数人の談話における盛り上がりを判定するために、笑いに関する 7 つの特徴量に着目している [4]：笑いの発生時間、笑いの大きさ、笑いの発生時間、同時に起きた笑いの人数、笑いの密度、発話一笑いの時間、笑いの人物情報。柴原は笑いの発生メカニズムを、哲学的視点、意味論的視点、精神分析的視点の 3 つの視点から検討している [5]。大島は、笑いの一種であるジョークの定義を述べている [6]。本研究では、マグロウの「無害・逸脱理論」および柴原の意味論的視点による笑いの定義を参考に、笑えるクエリ推薦の条件を検討した。

2.2 笑いを支援するシステム

笑いを支援するためのシステムの研究はいくつか行われている。Yang は文中のユーモアを文の違和感、曖昧さ、単語の善悪、主観的言葉の 4 つの素性で推定する手法と、ユーモアが含まれている文を品詞単位で評価することでユーモアを表現するために必要な語を特定する手法を提案している [7]。吉田は複数の言語資源をもとにして作成したユーモア含む対話システムを提案している [8]。入江は楽しみのある映像視聴をするために、韻律などの特徴をもとに映像中の笑いシーンを抽出する方法を提案している [9]。伏見は笑い声を再生することで被写体の自然な笑顔を撮影するアプリケーションを提案している

[10]. Tsujita は日常生活で笑う機会を増加させるために，笑うことで開く冷蔵庫を提案している [11]. 本研究では，ウェブ検索プロセス中に笑いの機会を作ることで笑いの支援を行う方法を提案する．

2.3 クエリ推薦

検索行動の効率化を目的としている既存のクエリ推薦とは目的が異なるクエリ推薦をする研究はいくつか行われている．佃や佐藤は，入力したクエリに関する情報をウィキペディアから取得し，ユーザに推薦することを提案している [12][13]．ウィキペディアの構造などから語を収集する点を，本研究でも取り入れている．また，佃の研究の意外な語の定義は，本研究の提案手法の意外性を決定する際に参考にした．

第 3 章

笑いを誘発するクエリの要件

本章では、笑いを誘発するクエリの変件について述べる．提案システムは、入力は 1 語のクエリ q とし、出力は q と組み合わせることで笑いを誘発する語の組み合わせのリストとする．例えば、入力として“アンパンマン”が与えられたとき、“アンパンマン-食中毒”など、入力クエリとある語の組み合わせのリストを推薦する．以下、意味論的視点で検討された笑いの発生メカニズムを参考に、提案システムの出力である 2 つの単語の組み合わせが満たすべき笑いの要件について検討する．

柴原によれば、「笑いが生まれるためには、テキストを構成している 2 つの異なるスクリプトが全般的ないし部分的にテキストに適合し、かつ両者の間に対立構造が存在する必要がある．」とされている [5]．柴原らの知見を 2 つの単語の組み合わせに適用した場合、「2 つの異なるスクリプトが全般的ないし部分的にテキストに適合し」の条件は、片方の語から片方の語の連想のしにくさ、すわわち語関係の意外性と言ひ表すことができる．語関係の意外性とは、ある語からある語を連想するとき、連想の経路が少ないことである．例えば、“アンパンマン- しょくぱんまん”の場合、「アンパンマン→アンパンマンのキャラクター→しょくぱんまん」や「アンパンマン→アンパン→食パン→しょくぱんまん」といった経路で連想される．一方、“アンパンマン- 食中毒”の場合、「アンパンマン→食べ物→食中毒」といった経路のみで連想される．複数の経路から連想できると、語関係は意外とはいひづらく笑ひも感じにくい．以上のように、本研究では語関係の意外性が笑ひを発生させる要件のひとつとする．

笑ひを誘発する語の組み合わせの要件の 2 つ目は、語の対立性である．柴原が主張する笑ひの条件における「両者の間に対立構造が存在する必要がある」の箇所は、語の対立性とみなすことができる．善と悪や生と死など、様々な観点から語の対立性が考えられるが、本研究では語の対立性を、2 つの単語が持つポジティブ・ネガティブといった印象の

違いと捉える．例えば，“アンパンマン”はポジティブなイメージを持つ語で，“食中毒”はネガティブなイメージを持つ語である．“アンパンマン”と“食中毒”は対立していることが分かる．意外性と対立性を組み合わせることで，より笑える語の組み合わせを検討することが可能となる．

第 4 章

提案内容

本章では，笑いを誘発するクエリの取得，ランキング方法について述べる．提案システムは以下の手順で笑いを誘発するクエリを推薦する：

1. クエリ q に対して組み合わせる関連語の候補集合 W_r を収集する．
2. q に対する関連語 $w_r \in W_r$ の意外性を求める．
3. q と関連語 w_r の対立性を求める．
4. 意外性と対立性を考慮して， q と関連語 w_r の組み合わせのランク付けを行う．
5. ランキング上位 k 件を笑いを誘発するクエリとして推薦する．

以下では各ステップについての詳細を述べる．

4.1 関連語の収集

本研究では，クエリの関連語を，笑いを誘発する語ペアを構成する語の候補とする．

関連語の候補は無数にありえるので，ランク付けの対象をある程度絞り込む必要がある．語の共起性などで候補語を絞り込むことも考えられるが，共起性と意外性は相反するので，共起性を用いて語の候補を集めようとするとは意外な語を取り逃してしまう可能性がある．そのため，本研究ではある語に対して意外な語および関連する語の両方を含む文書としてアンサイクロペディア^{*1}に着目した．アンサイクロペディアは物事を風刺的視点で解説する百科事典で，見出し語に関して関連するトピックを意外な観点から記述したコンテンツが多数収録されている．提案手法では，クエリを見出しとするアンサイクロペディ

^{*1} <https://ja.uncyclopedia.info/wiki>

ア記事に含まれている語のうち、アンサイクロペディアの他の記事で見出し語（記事名）として使用されている語をクエリの関連語として収集する．なお、関連語として見出し語のみに着目した理由は、一般的な語を排除するためである．

4.2 意外性

本研究では、語関係の意外性を求めるために、語の連想度と共起度に着目する．

語 w_x から語 w_y の連想度は、 w_x およびその同位語 w_r から w_y をどの程度連想できるかの度合いを表す．例えば“アンパンマン-食中毒”の場合，“アンパンマン”や“アンパンマン”の関連語であるアンパンマンの関連キャラクターからも“食中毒”を連想できないようであれば，“アンパンマン-食中毒”間の連想度は低いとみなせる．本研究では、連想度が低ければ意外性は低いと見なす．

語から語の連想度を求めるために、2つのステップを踏む．

1. クエリの同位語を求める．
2. クエリの関連語と同位語の関連語の連想度を求める．

クエリと同位語の類似度合と、クエリの関連語と同位語の関連語で共通している関連語の有無をもとに、クエリから関連語の連想度を求めている．本研究では、語 w_x と語 w_r が同じカテゴリに属している時、語 w_r は w_x の同位語であると定義する．また語 w_x と w_r の間で共通するカテゴリが多いほど、 w_r は w_x に対してより同位語らしいと、すわなち同位語度が高いとみなす．語 w_x に対する w_r の同位語度 $Syno(w_x, w_r)$ を以下の式で定義する：

$$Syno(w_r|w_x) = \frac{|Hype(w_x) \cap Hype(w_r)|}{|Hype(w_x)|} \quad (4.1)$$

ここで、 $Hype(w)$ は語 w が属するカテゴリ名の集合、 $|W|$ は語集合 W の要素数を表す．本研究では、上式の定義より $Syno(w_r|w_x) > 0$ となるとき、語 w_r は語 w_x の同位語であるとみなす．

同位語および同位語度のついて、図 4.1 を用いて具体的に説明する．ターゲット語として“アンパンマン”という語を考える．図が示すように、アンパンマンという語は「(アニメ) アンパンマンの登場人物」や「(人気アニメの) キャラクター」というカテゴリに属していると考えられる．一方、「(アニメ) アンパンマンの登場人物」としてはアンパンマンの他に「しょくぱんまん」、「(人気アニメの) キャラクター」としては「ドラえもん」などが考えられる．図 4.1 のように各語が属するカテゴリが得られている場合、例えば「ア

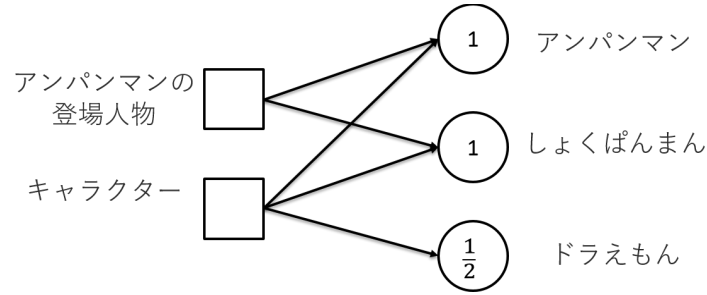


図 4.1 同位語度の計算イメージ

「アンパンマン」の上位カテゴリ語は 2 つあり、「ドラえもん」はそのうちの 1 つを共有しているため、アンパンマンに対するドラえもんの同位語度 $1/2$ となる（図中の数字はアンパンマンに対する語の同位語度を表す）。

提案手法を実装するにあたり、語が属するカテゴリの取得には、ALAGIN フォーラムから提供されている上位下位関係抽出ツール^{*2}を用いた。これはウィキペディアの記事の見出し語や階層構造などを用いて上位語と下位語を求めたデータであり、約 20 万の上位語と約 245 万の下位語を含む。本研究では、ある語 w の上位語を w が属するカテゴリと見なして、同位語度の計算を行った。

次に、同位語度を用いた語から語の連想度の計算方法について述べる。クエリ q に対するある語 w の連想度は、 q の関連語集合と q の同位語の関連語集合の共通要素の数で表現する。同位語の関連語は、クエリの関連語と同じようにアンサイクロペディアを用いた手法で収集する。連想度の計算では、同位語の同位語らしさも考慮する。これは、同位語が高いほどクエリとの関連性が強くなり、同位語から連想される語はクエリからも連想されやすくなると考えられるためである。以上の議論より、クエリ q に対する語 w の連想度 $Asso(w|q)$ の計算式は以下とした。

$$Asso(w|q) = \frac{1 + \sum_{w_s \in W_s(q) \cap Rel(w)} Syno(w|w_i)}{1 + |W_s(q)|} \quad (4.2)$$

ここで、 $W_s(q)$ はクエリ q の同位語の集合、 $Rel(w)$ は単語 w を関連語としてもつ語の集合を表す。

連想度の計算例を図 4.2 を用いて具体的に説明する。クエリ q として“アンパンマン”が与えられ、その同位語集合 $W_s(q)$ の要素として“しょくぱんまん”、“ドラえもん”の 2 語が得られたとする。今、“パン工場”が“アンパンマン”および“しょくぱんまん”の関連

^{*2} <https://alaginrc.nict.go.jp/hyponymy/index.html>

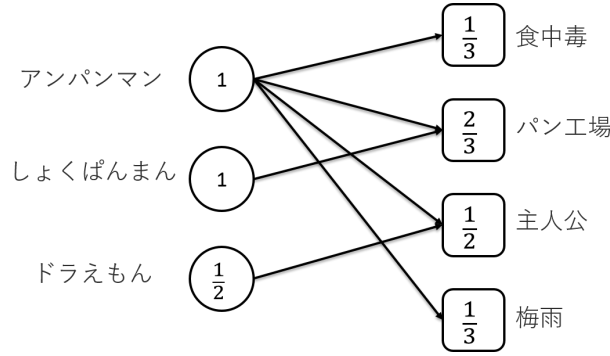


図 4.2 連想度の計算イメージ

語として得られたとする．図 2 の例の計算で “しょくぱんまん” の “アンパンマン” に対する同位語度 $Syno$ は 1 と得られているため， $Asso(\text{パン工場} | \text{アンパンマン})$ の値は， $(1+1)/(1+2)=2/3$ となる．すなわち，“アンパンマン” から “パン工場” の連想度は $2/3$ となる．

語と語の共起度は，2 つの語がウェブ文書中で共起する度合いとする．提案手法では，語 w_x と語 w_y の共起度 $corr(w_x|w_y)$ を以下の式で定義する：

$$Corr(w_x|w_y) = \frac{H(w_x \wedge w_y)}{H(w_x)} \quad (4.3)$$

ここで $H(w)$ は語 w でウェブ検索したときに得られるウェブページのヒット件数とする．上式の定義に基づき，提案システムはクエリ q とその関連語 w_r の共起度 $Corr(q|w_r)$ を計算する．なお，ウェブ検索結果のヒット件数の取得には Bing Web Search API^{*3}を用いた．

上記の連想度と共起度の定義に基づき，語と語の間の意外性を求める．クエリ q に対する関連語 w_r の意外性の計算式は下記の通りである．

$$Surp(q, w_r) = \frac{1}{\frac{Asso(w_r|q) + Corr(q|w_r)}{2}} \quad (4.4)$$

ステップ 1 で収集した関連語集合には，クエリに対して典型的な語，すわなち意外でない語も含まれている．先に定義した連想度と共起度を考慮することで，意外性の高い語が推薦ランキング上位に現れることが期待される．

^{*3} Bing Web Search API: <https://azure.microsoft.com/ja-jp/services/cognitive-services/bing-web-search-api/>

4.3 対立性

提案手法では語と語の対立性を求めるために、各語の極性に着目する。極性とは、語がポジティブな意味をもつかネガティブな意味をもつかを示す指標である。極性の値域は1から-1であり、ポジティブな語ほど1、ネガティブな語ほど-1に近づく。本研究では、単語 w_x と w_y 間の対立性 $Conf(w_x, w_y)$ を以下の式のように2つの語の極性の差として計算する：

$$Conf(w_x, w_y) = |pol(w_x) - pol(w_y)| \quad (4.5)$$

ここで $pol(w_x)$ は語 w_x の極性値を表す。語の極性値の取得は、高村らが開発した単語感情極性対応表を用いる [14]。

単語感情極性対応表には、提案手法の中で用いる語が掲載されていない場合もある。そこで、単語感情極性対応表に掲載されていない掲載語 w の極性を求めるために、ウェブ文書中で w と共起する語の極性値からの極性を推定する。以下に手順を示す。

1. 語 w でウェブ検索を行う。
2. 検索結果に含まれるウェブページのタイトルとスニペットから名詞を抽出する。
3. 抽出した名詞の極性の平均を取る。

ウェブ検索結果の取得には Bing Web Search API を用いた。取得する検索結果数は20件とした。名詞の極性の平均を取る際、名詞の出現頻度は考慮していない。

4.4 笑える語の組み合わせのランキング

上で定義した語間の意外性と対立性を用いて、クエリ q と q の関連語 w_r との組み合わせを笑いの誘発のしやすさ順にランキングを行う。

ランキング関数は以下の通りである。

$$Rank(w_r|q) = Conf(q, w_r) \cdot Surp(q, w_r) \quad (4.6)$$

第 5 章

評価実験

入力された語に対して，提案手法が笑いを誘発する語の組み合わせを提示できるかを評価するために評価実験を行った．実験では，提案手法および比較手法によって提示された語の組み合わせを人間が見て，実際に笑えるかどうかを調べることで提案手法の有効性を評価した．

5.1 比較手法

比較手法として，4 章で述べた語の間の意外性のみを考慮したランキング手法および語間の対立性のみを考慮した手法を用意した．

5.2 評価尺度

提案手法および比較手法が提示する語の組み合わせのランキングリストの評価指標として，適合率，平均適合率 (AP)，Mean Average Precision (MAP) を用いた．本稿で用いた適合率の定義は以下の通りである：

$$P@k = \frac{|Pair_s(k) \cap Pair_h|}{k} \quad (5.1)$$

ここで， $Pair_s(k)$ はシステムが返した語ペアのランキング結果の上位 k 件， $Pair_h$ は人間が面白いと判断した語ペア集合を表す．なお，本実験では評価者の半分以上が面白いと判断した語のペアを「面白い」と扱った．

MAP を求めるために AP を用いる．AP の定義は以下の通りである：

$$AP(q) = \frac{1}{|Pair_s \cap Pair_h|} \sum_{k=1}^{|Pair_s|} I(k)P@k \quad (5.2)$$

$pair_s$ はシステムが返したランキング結果リストを表す． $I(k)$ は第 k 位のペアが正解なら 1, 不正解なら 0 を返す関数を表す．MAP の定義は以下の通りである：

$$MAP(q) = \frac{\sum_{q \in Q} AP(q)}{|Q|} \quad (5.3)$$

ここで、 q はシステムに入力するクエリを表し、 Q は q の集合を表す．

$P@k$, AP , MAP は、値が大きいほど上位に面白い語を多くランキングできていることを示す．

5.3 評価用クエリ

実験では、表 5.1 に記した計 15 個の単語を評価用のクエリとした．評価用クエリはキャラクター、スポーツ、食べ物のカテゴリから 5 個ずつ用意した．人気キャラクターランキングといったランキングから、ウィキペディアとアンサイクロペディアの見出し語（記事名）として存在する語を選択した．

表 5.1 評価用クエリ一覧

カテゴリ	クエリ
キャラクター	アンパンマン, ミッキーマウス, きかんしゃトーマス, ムーミン, リラックマ
スポーツ	野球, サッカー, ボクシング, テニス, ゴルフ
食べ物	カレーライス, ラーメン, ハンバーグ, 肉じゃが, 炒飯

5.4 評価者

評価者の雇用のために、クラウドソーシングサービス「ランサーズ」^{*1}を用いた．ランサーズ内で 1 クエリのランキング評価を 1 タスクとして設定し、タスクを依頼した．1 タ

^{*1} <https://www.lancers.jp/>

スクにつき 6 人の評価者を募集した。タスクは合計 15 個あり，計 90 人のワーカーが評価に参加した。1 タスクあたり 56 円の報酬を支払った。

5.5 実験手順

実験の手順は以下の通りである。まず，表 5.1 の各クエリを提案手法と比較手法に入力し，出力された関連語を評価者に提示した。評価者には，以下のようなタスクに関する説明を提示した後，提示した関連語が面白いかどうかを，面白いと面白くないの 2 値で評価してもらった。

このタスクでは単語「アンパンマン」と別の単語の組み合わせが面白いかどうかを評価していただきます。以下にいくつかの単語が提示されるので，単語「アンパンマン」との組み合わせが「面白い」か「面白くない」かを評価してください。2 つの単語からイメージできることを面白いと感じたら「面白い」，面白くないと感じたら「面白くない」を選択してください。

例えば単語「交通事故」と単語「バナナ」の組み合わせで，バナナが原因で自動車が滑ってしまい交通事故が起きたことをイメージし，それが現実的にありえないので面白いと感じたら「面白い」を選択してください。

第 6 章

結果

$P@5$, $P@10$, $P@15$, AP , MAP の評価結果を表 6.1 に記す. 表 6.1 の MAP の値は, 「意外性のみ」, 「提案手法」, 「対立性のみ」の順で大きかった (意外性のみ : 0.288, 提案手法 : 0.247, 対立性のみ : 0.222). このことから, 意外性のみを考慮する手法が平均的には笑いを誘発するクエリを推薦できることが分かる.

一方で, 表 6.1 の AP の値から, クエリによっては提案手法が最も優れた結果を出力していることが分かる. 例えば, “リラックマ ($AP=0.313$)” や “テニス ($AP=0.315$)” や “ラーメン ($AP=0.380$)” などの 6 つのクエリで提案手法は優れた結果を出した. また, 提案手法では優れた結果を出せなかったクエリがあることも分かる. 例えば, クエリが “きかんしゃトーマス” だった場合, 提案手法は 3 つの手法の中で最も低い値を出した ($AP=0.243$). MAP では「対立性のみ」の手法が最も低い結果を出していたが, AP を見ると「対立性のみ」を用いた手法が他の手法より優れた結果を出したクエリがあることが分かる. 例えば, “ミッキーマウス ($AP=0.364$)” や “サッカー ($AP=0.250$)” や “カレーライス ($AP=0.100$)” などの 6 つのクエリで対立性のみを用いる手法は優れた結果を出した. 以上から, クエリ毎に適している手法があることが分かる.

表 6.1 $P@5$, $P@10$, $P@15$, AP , MAP の評価結果

クエリ名	p@5			p@10			p@15			AP		
	提案手法	意外性のみ	対立性のみ	提案手法	意外性のみ	対立性のみ	提案手法	意外性のみ	対立性のみ	提案手法	意外性のみ	対立性のみ
アンパンマン	0.6	0.6	0.6	0.5	0.5	0.5	0.467	0.533	0.467	0.543	0.546	0.513
ミッキーマウス	0.2	0.2	0.2	0.2	0.3	0.4	0.200	0.200	0.400	0.266	0.272	0.364
きかんしゃトーマス	0	0.4	0	0.2	0.2	0.2	0.267	0.133	0.333	0.243	0.253	0.313
ムーミン	0.4	0.2	0.2	0.2	0.2	0.2	0.200	0.200	0.200	0.424	0.443	0.355
リラックマ	0.2	0.2	0.2	0.3	0.1	0.2	0.267	0.200	0.267	0.313	0.228	0.290
野球	0.2	0	0	0.1	0.1	0.1	0.067	0.067	0.067	0.159	0.137	0.121
サッカー	0.2	0	0.2	0.2	0.1	0.2	0.133	0.067	0.133	0.236	0.118	0.250
ボクシング	0	0	0.2	0.2	0.1	0.2	0.133	0.133	0.133	0.171	0.143	0.200
テニス	0.4	0.2	0.2	0.2	0.2	0.2	0.200	0.200	0.200	0.315	0.236	0.249
ゴルフ	0	0.2	0	0.1	0.1	0	0.067	0.067	0.067	0.125	1	0.091
カレーライス	0	0	0.2	0.1	0	0.1	0.067	0	0.067	0.086	0.058	0.100
ラーメン	0.6	0.4	0	0.4	0.4	0	0.333	0.333	0.067	0.380	0.321	0.132
ハンバーグ	0	0	0	0.2	0	0.2	0.133	0	0.133	0.162	0.135	0.162
肉じゃが	0.2	0.2	0	0.1	0.1	0	0.067	0.133	0.067	0.163	0.321	0.073
炒飯	0	0	0	0.1	0.1	0	0.133	0.067	0.067	0.119	0.111	0.112
MAP										0.247	0.288	0.222

各クエリの出力結果の上位 20 件の適合率の推移を表すグラフを図 6.1 から図 6.15 に記す。結果が 20 件に満たないものは結果の上限までをグラフに記す。これらのグラフの $P@k$ が上昇している k の値を見ることで、笑える語が出現した順位が分かる。また、その順位を見ることでクエリに適した手法が分かる。例えば、図 6.12 から、クエリを“ラーメン”とした場合には、提案手法を用いると上位に笑える語を出現させることができると分かる。

提案手法の評価結果が良かったクエリと悪かったクエリの出力結果上位 5 件を表 6.2 に記す。表 6.2 中で、語の後ろに○がついている語は笑えると判定された語で、×がついている語は笑えないと判定された語である。

表 6.2 提案手法の出力結果の上位 5 件 (○は笑えると判定された組み合わせ, ×は笑えないと判定された組み合わせ)

順位	良い結果		悪い結果	
	アンパンマン	ムーミン	カレーライス	炒飯
1	脅迫○	名古屋共和国×	聖人×	笑い飯×
2	アソパソマソ○	ピクミン○	薬師如来×	名取羽美×
3	ワンパンマン○	カバ×	捨身飼虎×	眠気×
4	段ボール肉まん×	桶狭間○	両生類×	クモ×
5	常識的に考えて×	サンタクロース×	徳×	タバスコ×

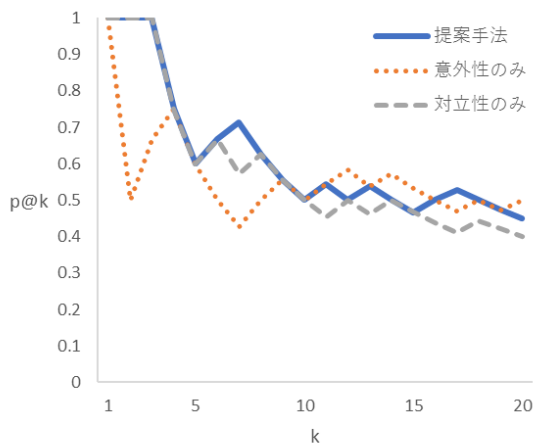


図 6.1 アンパンマンの適合率の推移

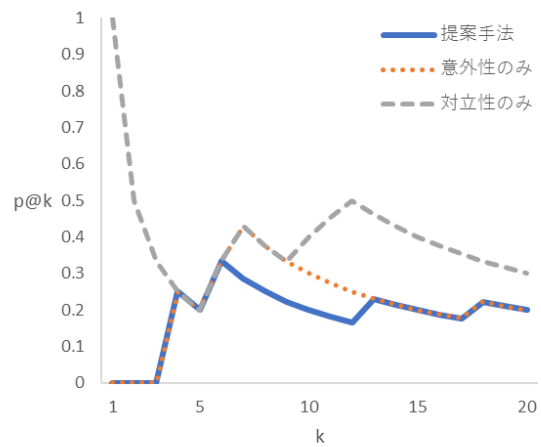


図 6.2 ミッキーマウスの適合率の推移

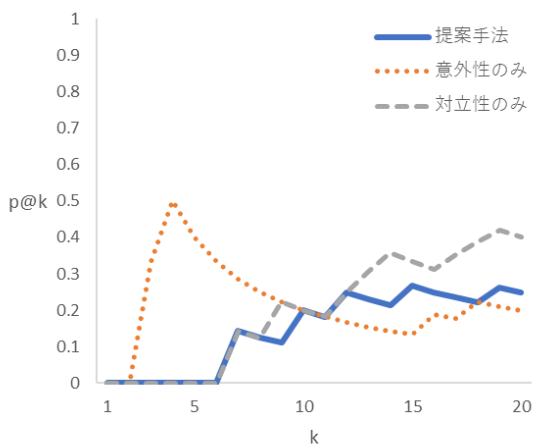


図 6.3 きかんしゃトーマスの適合率の推移

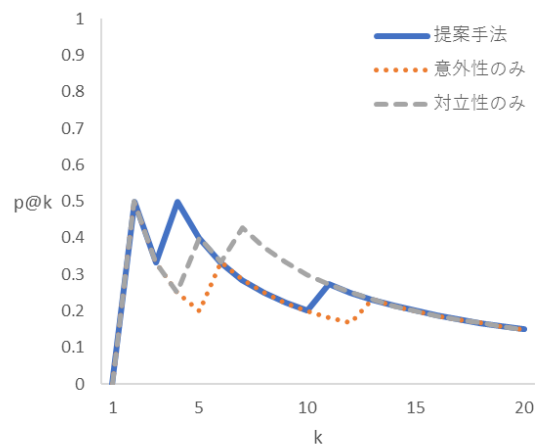


図 6.4 ムーミンの適合率の推移

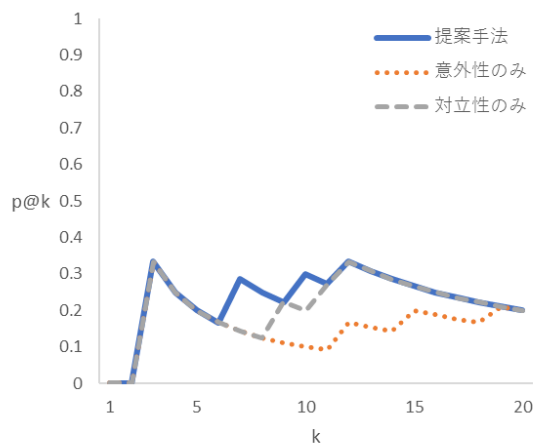


図 6.5 リラックマの適合率の推移

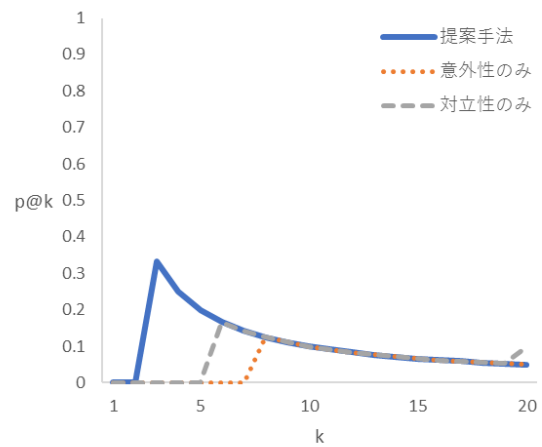


図 6.6 野球の適合率の推移

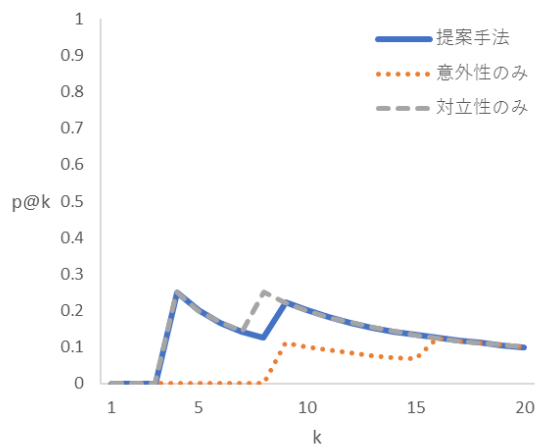


図 6.7 サッカーの適合率の推移

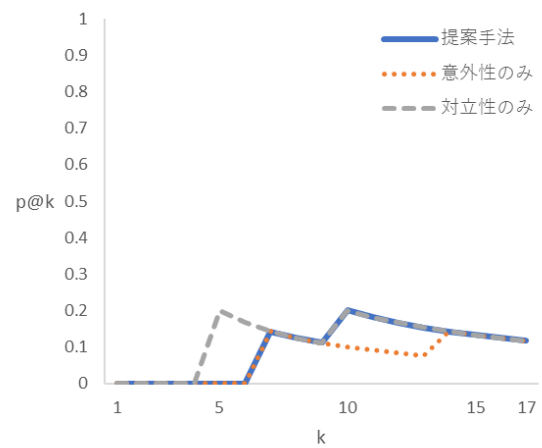


図 6.8 ボクシングの適合率の推移

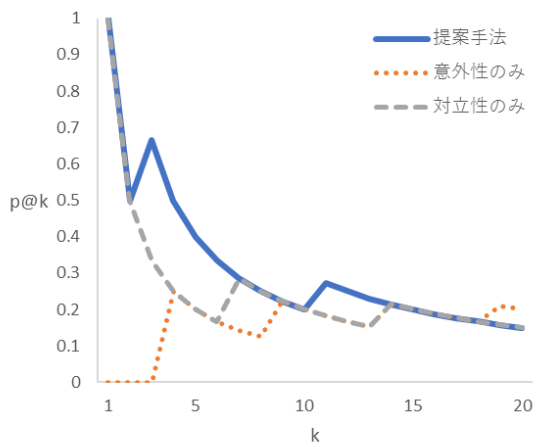


図 6.9 テニスの適合率の推移

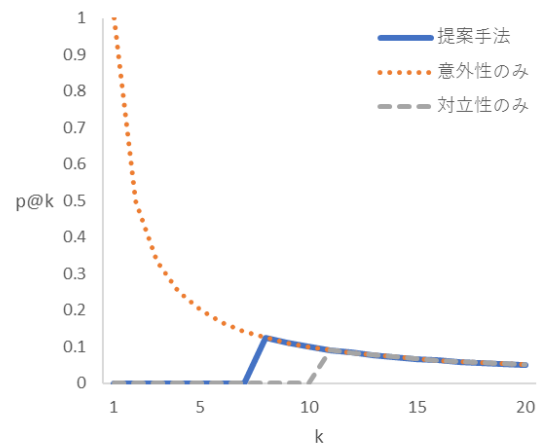


図 6.10 ゴルフの適合率の推移

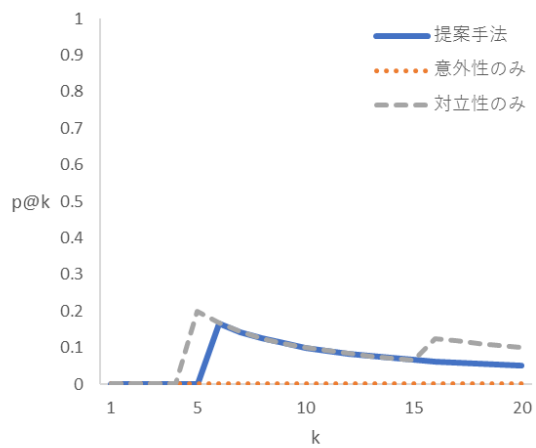


図 6.11 カレーライスの適合率の推移

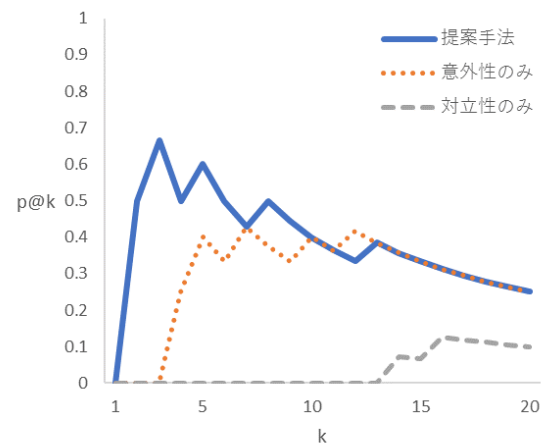


図 6.12 ラーメンの適合率の推移

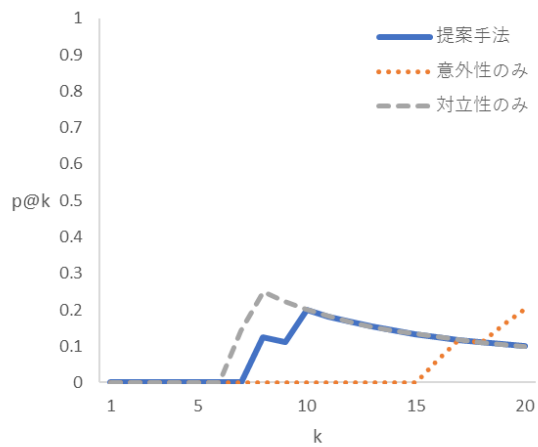


図 6.13 ハンバーグの適合率の推移

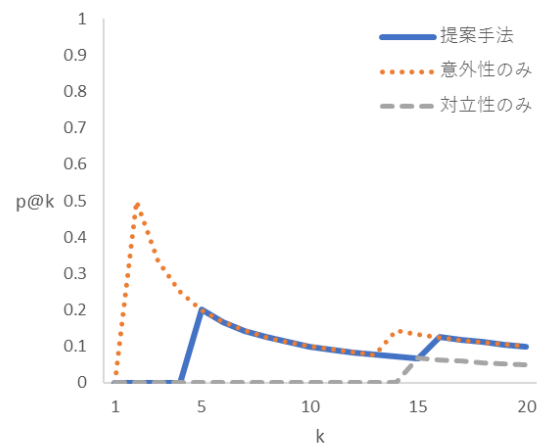


図 6.14 肉じゃがの適合率の推移

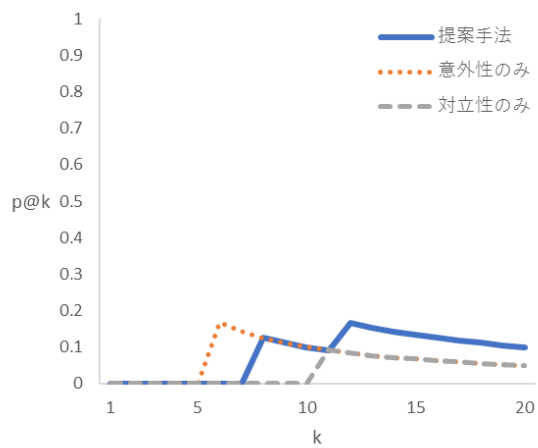


図 6.15 炒飯の適合率の推移

第 7 章

考察

表 6.1 の *MAP* の評価結果から，意外性のみを用いた手法が平均的に優れていることが分かるが，*AP* の値を見ると，クエリによっては提案手法や対立性のみを用いた手法が優れている場合があった．クエリ毎に優れている手法が異なった原因として，3つの原因が考えられる．

1. 意外性と対立性を用いることが笑いの条件として適切ではない．
2. 意外性と対立性の考慮の方法が適切ではない．
3. 提案手法のランキング関数が適切ではない．

1つ目の原因は，提案手法で用いた笑いを発生させるための条件が適切ではなかったことである．これについては，マグロウの笑いの理論 [3] や柴原の笑い発生のメカニズム [5] から，意外性と対立性が笑い発生の条件の一つであることから，1つ目の項目が原因であることは考えにくい．

2つ目の原因は，意外性と対立性を考慮する際の計算が上手く機能していないことである．意外性は連想度と共起度を用いている．連想度と共起度は両方ともウェブページの情報に依存する求め方をしているので，語の意外性を誤って算出してしまう場合が考えられる．対立性の求める方法に，語と語の極性の差を用いている．各語の極性を求めるためにウェブ検索の結果を使用しているので，極性は検索結果に依存してしまう．ポジティブな意味の語だとしても，ネガティブな語として扱ってしまう場合がある．極性の差を取る方法だと，極性の異なる組み合わせでも同一の対立性のものとみなしてしまう場合が考えられる．例えば，極性が1と0の単語の組み合わせと-1と0の単語の組み合わせの場合，どちらも極性の差は1となり，同一のものとして扱ってしまう．笑いの条件を考慮する方法によって提案手法が機能しなかったことが考えられる．

3つ目の原因は，提案手法のランキング関数が不適切だったことである．提案手法のランキング関数は，意外性と対立性の積をとっている．対立性は極性の差をとっていることから，0となる場合がある．その場合，意外性の値を考慮することができない．

以上の考察から，クエリによって提案手法が機能する場合としない場合がある原因は，意外性と対立性の計算方法および提案手法のランキング関数にあると考える．

第 8 章

おわりに

本稿では，ウェブ検索プロセスにおいて笑いの機会を提供するために，笑いを誘発するクエリの推薦技術についての提案を行った．提案手法を用いた場合，“アンパンマン”というクエリを入力すると，笑いの発生メカニズムである適合と対立構造に基づいて推定された”食中毒”といったクエリを返す．提案手法で，これまでのクエリ推薦にはない笑いを誘うクエリをユーザに提示し，検索するたびに笑いの機会を与えることができると期待した．

実験結果から，意外性のみを用いた手法に劣るものの，意外性と対立性を考慮することで笑いを誘発するクエリを推薦できることを明らかにした．本研究で得られた知見を用いることで，日々のウェブ検索を行う際に笑いを発生させることが期待できる．一方で，提案手法が比較手法に劣ってしまうという大きな課題が残っている．この課題は，対立性の考慮の仕方や語の収集する情報源を増やすことで改善を試みる．また，図 1.1 のようなユーザインターフェースで，実際に人は笑えるのかを評価することも今後の課題である．

本研究では，笑いを誘発するクエリの要件として意外性と対立性を用いた．これらの要件を満たす笑いが，笑いの全てではない．笑いを誘発する要件は他にもあるので，それらも考慮してよりよい手法を提案することを考えている．

参考文献

- [1] 総務省 (2017) 「情報通信メディアの利用時間と情報行動に関する調査」.
- [2] 大原哲也 (2018) 「笑いの抗加齢効果」, いのち輝く未来社会をめざすビジョン推進のための「10 歳若返り」ワークショップ.
- [3] Peter McGrow, Joel Warner(2015) 「世界笑いのツボ探し」, CCC メディアハウス.
- [4] 島田和孝, 楠本章裕, 横山貴彦, 遠藤勉 (2012) 「複数人談話おける笑いの情報を考慮した盛り上がり判定」, 『電子情報通信学会技術研究報告』112(110), pp.25-30, 一般社団法人電子情報通信学会.
- [5] 柴原直樹 (2006) 「笑いの発生メカニズム」, 『近畿福祉大学紀要』7(1), pp.1-11, 近畿医療福祉大学.
- [6] 大島希巳江 (2005) 「高コンテキスト社会と低コンテキスト社会のコミュニケーションにおけるユーモア」, 『笑い学研究』12(0), pp.29-39, 日本笑い学会.
- [7] Diyi Yang, Alon Lavie, Chris Dyer, Eduard Hovy, “Humor Recognition and Humor Anchor Extraction”, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp.2367-2376, 2015.
- [8] 吉田裕介, 萩原将文 (2014) 「複数の言語資源を用いたユーモアを含む対話システム」, 『日本知能情報ファジィ学会誌』26(2), pp.627-636.
- [9] 入江豪, 日高浩太, 宮下直也, 佐藤隆, 谷口行信 (2008) 「個人撮影映像を対象とした映像速覧のための “笑いシーン” 検出」, 『映像情報メディア学会誌』62(2), pp.227-233.
- [10] 伏見遼平, 福嶋政期, 苗村健 (2014) 「笑い声呈示により自然な笑顔を撮影するカメラの提案」, 『エンタテインメントコンピューティングシンポジウム 2014 論文集』, pp.26-31.
- [11] Hitomi Tsujita, Jun Rekimoto, “Smiling Makes Us Happier: Enhancing Positive Mood and Communication with Smile-Encouraging Digital Appliances”, *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp*

2011), pp.1-10, 2011.

- [12] 佃洗撰, 大島裕明, 山本光穂, 岩崎弘利, 田中克己 (2014)「語の認知度と語間の関係の非典型度に基づく Wikipedia からの意外な情報の発見」,『情報処理学会論文誌データベース』7(1), pp.1-17.
- [13] 佐藤進也 (2018)「検索対象の多面的理解支援のための Wikipedia 記事中の列挙を利用した関連情報発見」,『知能と情報 (日本知能情報ファジィ学会誌)』30(6), pp.788-795.
- [14] 高村大也, 乾孝司, 奥村学 (2006)「スピンモデルによる単語の感情極性抽出」,『情報処理学会論文誌ジャーナル』47(2), pp.627-637.

謝辞

本研究の遂行ならびに論文の作成にあたり，ご指導を賜りました静岡大学情報学部講師 山本祐輔先生に謹んで深謝の意を表します．本研究を遂行するうえで，副査として有益な御助言を賜りました静岡大学情報学部准教授 福田直樹先生に心より感謝の意を表します．本研究の遂行ならびに研究室活動にご協力いただいた，静岡大学山本研究室の皆様には感謝いたします．最後に，これまで暖かく見守ってくれた両親と兄に感謝します．

2019 年 3 月 梅田 浩郎