人間とコンピュータが協調するための
評価指標マネジメントインターフェースの
実現に関する研究

２０２０年　３月

松本　恭平

人間とコンピュータが協調するための
評価指標マネジメントインターフェースの
実現に関する研究

松本　恭平

システム情報工学研究科

筑波大学

２０２０年　３月

博士論文概要

本論では，人間とコンピュータが協調するためのインターフェースである評価指標マネジメントインターフェースの提案を行う．また，その具体的実現例として，コンピュータ上に計算可能なシステムとして実装された複数のインターフェースを示す．

人間は社会での活動において，例えば企業内における経営判断や施策立案など，様々な場面において意思決定を行う．これまでは，企業の経営者や，もっと小さいレベルでは上司などといった存在が，彼らのこれまでの経験やそれに基づく明文化されない知識，平たく言えば勘やコツといったものによって，主観的な側面をもって意思決定を行ってきた．その一方で，世の中のあらゆるものにコンピュータを介したシステムが導入され，人間の様々な活動をセンサーによって感知することが可能となった現代においては，人間の活動をデータ化し，そのデータを分析することによって，データを根拠とした意思決定が可能となってきている．こうした潮流は，集約されたデータを基にアクションを決定することから，データインテンシブ (Data-intensive, データ集約型) と呼ばれている．

このような背景を基に，人間がコンピュータの力を借りてデータ分析を行うための基盤として，人間とコンピュータが協調するための評価指標マネジメントインターフェースを提案する．ここでの評価指標とは，例えば画像分類問題では分類精度，将棋や囲碁の強化学習では最終的な勝敗といったように，コンピュータ上で計算可能で，その計算結果に基づいてモデルの最適化を行える指標を指す．このインターフェースは人間に対して評価指標の選択を支援することによって，意思決定を支援する．これによって，人間はコンピュータの力を借りて，自身の知識の検証を行うことや，コンピュータから新たな知見を得ることが可能となる．

このようなインターフェースを利用して人間がコンピュータの力を借りるためには，人間側の要望をコンピュータに正しく伝える手法が必要となる．そのため本論においては，要望に合わせて評価指標を選択するマネジメントインターフェースを実現するとともに，人間側の要望をコンピュータに伝えるための手法の提案や，新たな評価指標の提案，コンピュータの計算結果の解釈を支援する手法の提案を行う．また，上記各手法の具体的実現例について述べる．

目 次

図 目 次

第1章　序論

人間の社会での活動において，意思決定は重要なファクターである．例えば，企業の経営者は，経営状況や世相を見て経営判断を行う．もっと小さなレベルで言えば，会社における上司やプロジェクトリーダーの指示や，ひいては各個人が行動を選択する行為も意思決定である．これらは，意思決定を行おうとする人間が，これまでの経験から得た知識，該当分野の専門家の知識，さらには各個人の勘・コツといったものよって，主観的に判断されてきた．

その一方で昨今では，人間のあらゆる行動がセンサーで感知可能になり，人間の行動をデータとして取得することが可能となった．そのため，取得したデータに意思決定の根拠を求める潮流が生まれている．こうした潮流は，集約されたデータを基にアクションを決定することから，データインテンシブ (Data-intensive, データ集約型) と形容されている．

Tony ら [1] はこのデータインテンシブに行われる科学を第四のパラダイムと評している．まず初めに存在した科学は現実の事象を数学的あるいは経験的な手法によって説明する科学であり，これが第一のパラダイムである．その後，ケプラーの法則，ニュートン力学やマクスウェルの方程式に代表されるような様々な理論を構築して科学を説明するようになったのが第二のパラダイムである．しかし，理論構築による説明も万能ではなく，複雑に条件が絡み合う問題に対応できないことがわかってくると，今度は現実の事象をシミュレーションして解き明かそうとした．これが第三のパラダイムである．そして今，現実の事象もシミュレーションも多様で巨大なデータを生み出してコンピュータに蓄積され続けている．この巨大なデータから関係性を見出す科学こそが第四のパラダイムというわけである．このような議論を裏付けるように，近年における機械学習の発展は目覚ましい成果を挙げている．これは，これまで人間がデータ間の関係性を発見して行われてきた事象のモデル化が，機械学習によって自動的に行えるようになったからだと考えている．

このような背景において重要なことは，人間とコンピュータが適切に役割分担を行い，データ分析を行う枠組みであると考えられる．この適切な役割分担によって，各個人の能力をコンピュータがエンハンス (促進，増進) することを，本稿では人間とコンピュータの協調と呼ぶ．コンピュータによるデータ間の自動的な関係性の発見が可能となった今，人間が考えなければならないことは，どのような知識をデータから発見したいか，そのためにはどのようなデータを使わなければならないのか，になりつつある．この人間の知識創造をコンピュータが促し，さらにはコンピュータが知識を検証し，新たな知見を授けるといったことまで行えるような枠組みを目指す．その枠組みとして本稿では，人間がコンピュータと協調してデータ分析を行うための，人間とコンピュータの間を取り持つインターフェースである評価指標マネジメントインターフェースを提案する．

我々は元来，コンピュータ上への専門知の統合・集約について研究を行ってきた．各研究分野の専門家が一生をかけて研究した内容をコンピュータ上に計算可能な形で実装し，それをさらに言葉というメタレベルで連結することにより，専門知の統合・集約を行おうというものである．このような統合・集約によって，コンピュータの役割は単なる計算・シミュレーションから，人間一人では生み出すことのできない新たな知識の提案や創造になる．このような統合・集約は，メディアコンテンツが持つ意味的な情報を言葉のメタデータとして抽出する枠組みである “Media-lexicon Transformation Operator”(以下 $\mathcal{ML}$ と表記)[2] と，データ間の意味的，感性的な等価性，類似性，関連性を文脈に応じて動的に計量するモデルである “意味の数学モデル”[3] によって実現されてきた．

$\mathcal{ML}$ は，様々なメディアコンテンツを対象とし，そのメディアコンテンツを表す意味的な言葉を重み付き単語群のメタデータとして抽出する作用素であり，次のように表される．

$$
\begin{aligned}
\mathcal{ML}(Md) &: Md \mapsto Ws, \\
\text{ただし，} & \\
\mathcal{ML} &: \text{Media-Lexicon transformation operator}, \\
Md &: \text{メディアコンテンツ}, \\
Ws &: \text{重み付き単語群}.
\end{aligned}
$$

また，$\mathcal{ML}$ の逆演算，すなわち重み付き単語群からメディアコンテンツを生成する作用素についても考えることができ，特に統計的情報を用いた制約条件によってメディアコンテンツを生成する作用素が，統計的一般化逆作用素 Stochastic Generalized Inverse Media-lexicon Transformation Operator($i\mathcal{ML}$)[4] として提案されている．$\mathcal{ML}$ および $i\mathcal{ML}$ の概要を図 1.1 に示す．この $\mathcal{ML}$ の枠組みを用いて，例えば楽曲データから印象語群を取り出す $\mathcal{ML}$[5] や，その逆となる印象語群から楽曲を自動生成する $i\mathcal{ML}$[6]，画像データからその印象語群を取り出す $\mathcal{ML}$[7] などが提案されている．

意味の数学モデルは清木，北川らによって考案された文脈依存の意味的連想記憶モデルである [3]．このモデルの根底にあるのは，辞書のような人間の知識や英知が集約された集合を用いて形成された意味空間である．この意味空間から文脈に対応した部分空間を選択することによって，データがもつ意味的あるいは感性的な同一性，類似性，関連性を文脈や状況に応じて動的に計量することが可能となる．ここでいう “ある概念” とは言葉だけにとどまらず，$\mathcal{ML}$ との組み合わせによって，例えば画像データと楽曲データの意味的な近さを計量することが可能となる．

意味の数学モデルにおいて，文脈に応じて部分空間を選択し計量を行う操作の概念図を図 1.2 に示す．意味空間に配置された言葉のベクトルを，文脈に応じて選択された部分空間に意味的に射影することによって，言葉のベクトルの動的な変化を表現することが可能となる．

さらにこれらを一般化し，ビッグデータ分析における知識創造・知識利活用に適用可能なモデルとして岡田らが提案したのが知識創造サイクルモデルである [4]．概念図を図 1.3 に示す．知識創造と知識利活用を互いに逆の演算とみなしてサイクルを構成することにより，知識の検証が可能であることを示している．

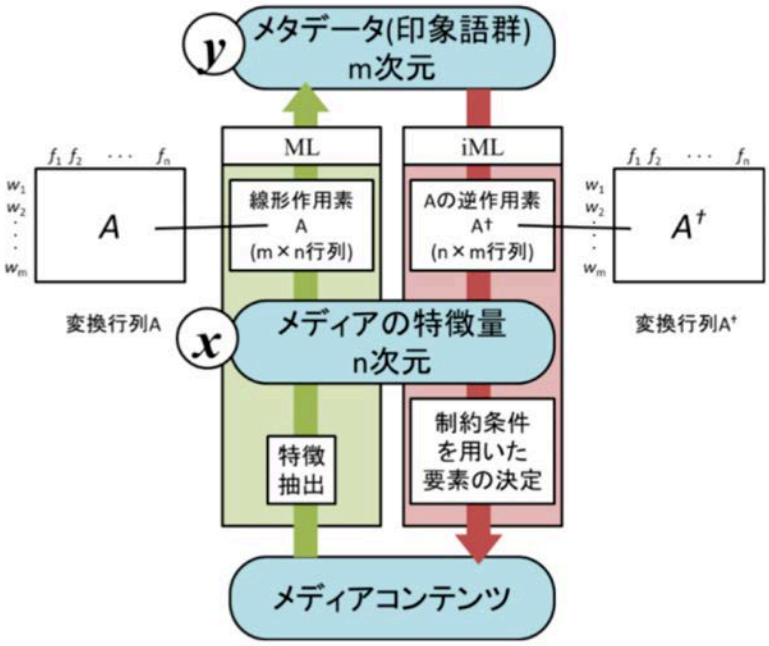


図 1.1: Media-lexicon Transformation Operator($\mathcal{ML}$) とその逆作用素 Stochastic Generalized Inverse Media-lexicon Transformation Operator($i\mathcal{ML}$)

我々はこれらの提案されたモデルを基に，評価指標マネジメントインターフェースを実現する．ここでの評価指標とは，例えば画像分類問題では分類精度，将棋や囲碁の強化学習では最終的な勝敗といったように，コンピュータ上で計算可能で，その計算結果に基づいてモデルの最適化を行える指標を指す．この指標は，人間がデータ分析をして行いたいこと，すなわち要望に対応する．評価指標マネジメントインターフェースは，人間側の要望を適切に理解し，その要望に対応した関係性の発見が行えるように，適切に整形したデータと選択した評価指標をコンピュータに渡す．また，コンピュータが発見した関係性を，人間が理解可能な形で表現して人間に示す．これらの実現のために，人間側の要望をコンピュータに伝えるための手法の提案や，新たな評価指標の提案，コンピュータの計算結果の解釈を支援する手法の提案など，インターフェースを実現する要素についても述べる．また，提案した枠組みである評価指標マネジメントインターフェースの具体的実現例を複数述べる．

本稿では，第 2 章にて評価指標マネジメントインターフェースの概要を述べる．その具体的実現例として，第 3 章では画像特徴手法自動選択メタシステム，第 4 章では授業内発想支援システム *AI-Josyu*，第 5 章では CM 字コンテ好感度予測システム *CREATIVE BRAIN*，第 6 章では Web アクセスログからのデモグラフィックデータ予測，第 7 章ではオウンドメディアにおけるカスタマージャーニー分析を述べ，第 8 章で総括する．

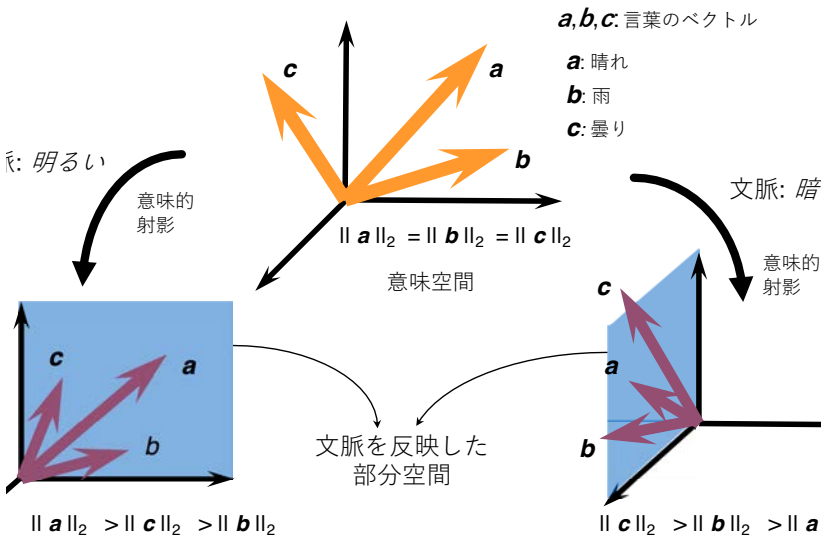


図 1.2: 意味の数学モデルにおける部分空間の選択と意味的射影

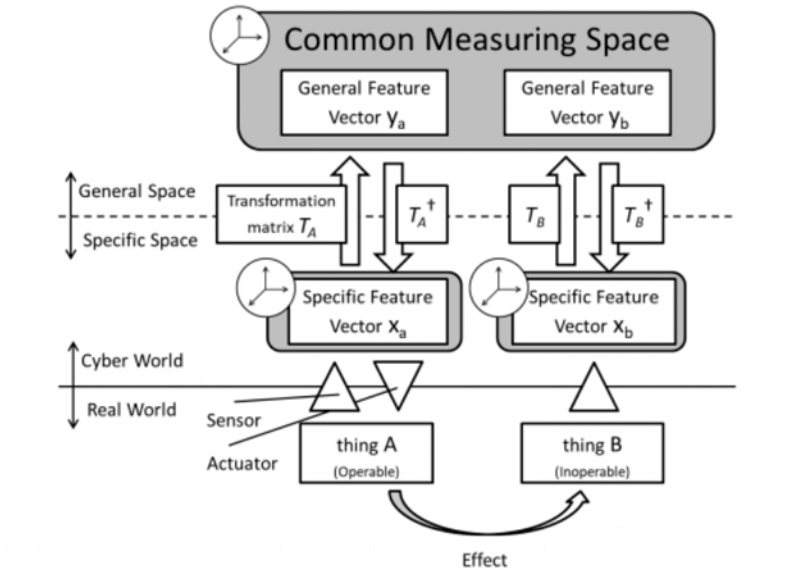


図 1.3: 知識創造サイクルモデルの概念図 (複合分野)

第2章　評価指標マネジメントインターフェース

この章では，人間がコンピュータと協調してデータ分析を行うための，人間とコンピュータの間を取り持つインターフェースである評価指標マネジメントインターフェースの概要を述べる．

図 2.1 に，提案する評価指標マネジメントインターフェースに求められることを示す．ここではコンピュータによって明らかにしたいことを要望と呼ぶ．この要望をコンピュータが理解可能な形で伝える必要がある．そのためにはコンピュータに使用可能なデータがどれだけあるかと，モデルを最適化するための評価指標が必要となる．また，コンピュータがデータと評価指標を基に生成したモデルと，そのモデルによる計算結果を人間が解釈可能な形で表現されている必要があると考えられる．

図 2.2 は，人間とコンピュータの協調における，要望に応じた入力データと評価指標の再定義サイクルを示した図である．図中の数字は，それぞれの要素が行われる順番を示す．それぞれの要素が行う内容について，以下に記述する．

1. ユーザやクライアントの要望
   コンピュータを使用して事象をモデル化するにあたり，そのモデルによって行いたいことがが要望である．コンピュータを操作する技術者は，この要望を基にモデル生成を行わせる．

2. コンピュータに与えるデータと評価指標の定義
   要望を達成できるように，モデル化に必要なデータと，コンピュータ上で計算可能な評価指標を技術者 (あるいは研究者など) が定義してコンピュータに渡す．

3. メタモデルによるモデル化
   入力されたデータと評価指標を基に，コンピュータがモデルを生成する．

4. 生成されたモデルによる結果の表示
   生成されたモデルを使用して得られた結果を技術者に提示する．

5. 要望と得られた結果の比較
   得られた結果が要望を満たしているかどうか，比較して判断する．

6. データと評価指標の再定義・修正
   得られた結果が要望を満たすように入力するデータや評価指標を吟味し，適宜再定義・修正を行う．

図 2.1: 評価指標マネジメントインターフェースに求められること

このようなサイクルを回すことにより，要望を満たすデータと評価指標をコンピュータと協調しながら決定していくことが可能となる．このサイクルを実現するにあたっては，要望と評価指標，あるいは入力するデータと使用できるデータの間にギャップが存在する．また，生成されたモデルから得られた結果と要望の間にもギャップが存在する．そのギャップを埋めるためにコンピュータと人間の仲立ちする役割を持つのが評価指標マネジメントインターフェースである．

インターフェースの概要図を図 2.3 に示す．評価指標マネジメントインターフェースは人間とコンピュータの間に立って，それぞれが連携することによってサイクルを構成する．

また，このインターフェースは図 2.4 のように，要望を依頼するクライアントや，インターフェースを使用するユーザという存在を考慮し，人間側が 2 つになる場合もある．この場合，サイクルを構成するために，人間同士による話し合いなどのすり合わせが必要となる．実システムを伴うような実現例の場合はたいていこのように表される (例えば，第 4 章 授業内発想支援システム，第 5 章 CM 字コンテ好感度予測システムの場合など．)．

図 2.5 は，本論文における各章の内容とインターフェースのそれぞれの要素の対応付けを行ったものである．すべての要素を実現したもの，一部の要素を実現したものが存在するため，要素ごとについて対応する内容を記述している．このサイクルを実現するために，人間・コンピュータ・インターフェースのそれぞれの間に必要な要素を以下に述べる．

2.1 人間 (技術者) からインターフェース

人間がデータ分析によって何を明らかにしたいかを伝える部分である．インターフェースは複数の評価指標をマネジメントし，要望に応じた評価指標を選択する．

この実現のために，インターフェースへのデータの入力方法と，要望の伝達手法を整備する必要がある．インターフェースそのものは使用するデータの取捨選択を行うことができな

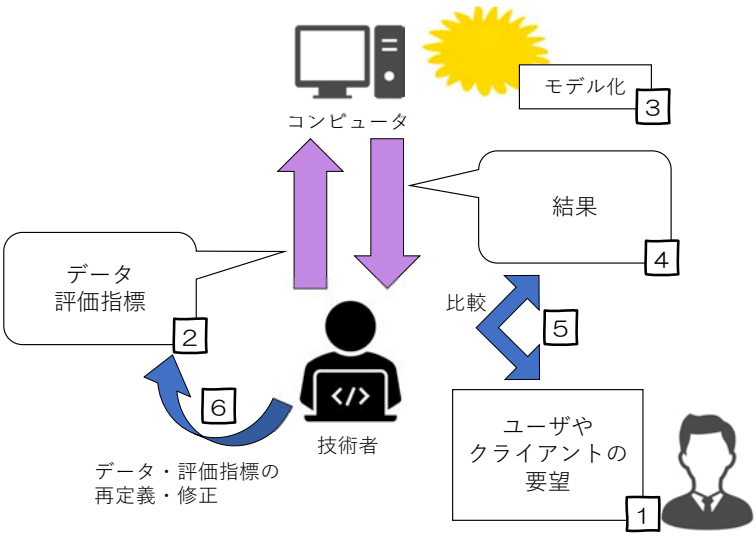


図 2.2: 要望に応じた入力データと評価指標の再定義サイクル

いため，人間が必要に応じてデータを整形して入力する必要がある．また，要望に関しても，コンピュータが理解可能な形である評価指標で表現された形での伝達が必要となる．

本稿においては要望に応じたデータの入力方法の整備という点で第 3 章 画像特徴量手法自動選択メタシステム，第 4 章 授業内発想支援システム，第 5 章 CM 字コンテ好感度予測システムがこの部分に該当する．また，要望の伝達のための評価指標の提案として第 7 章 カスタマージャーニー分析が該当する．

2.2 インターフェースとコンピュータ

インターフェースは要望である評価指標と適切に整形されたデータをコンピュータに渡す．コンピュータは，ニューラルネットワークを代表とする関係性を自動で発見可能なメタモデルを用いて，評価指標を最適化するモデルを自動選択する．

また，コンピュータリソースに応じたモデルの選択を行う．複数のパラメータを取りうる問題に対して総当たりの探索を行うことはできず，そもそも連続値をパラメータとして取り扱う場合には探索範囲が無限に分解可能となってしまい，最良のモデルを選択することは実質的に不可能となってしまう．問題に対して最良となるモデルでなくとも，要望が達成され

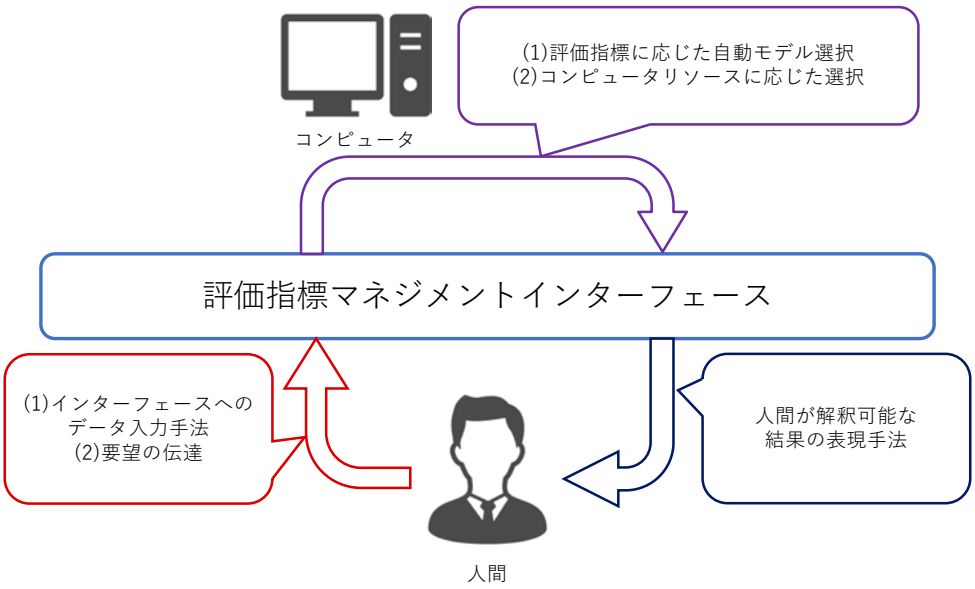


図 2.3: 評価指標マネジメントインターフェースの概要図

る範囲で，実時間で探索を行える枠組みを整備する．

第 3 章 画像特徴量手法自動選択メタシステムは画像の専門知を組み合わせてモデルそのものを構築するメタモデルであり，この部分に該当する．また，複数モデルの比較を行う観点から，第 6 章 デモグラフィックデータ予測が該当する．

2.3 インターフェースから人間 (ユーザ)

インターフェースはコンピュータから受け取った結果を，人間が解釈可能な形に表現して伝える．人間はこのインターフェースから与えられた結果を見て，要望が正しかったか吟味し，必要に応じて要望を修正し，再度インターフェースに要望を伝える．このインターフェースの実現により，人間は収集したデータから見出したい関係性を評価指標に応じて発見することができる．また，サイクルの形で実現することにより，知識創造サイクルモデルで提案されている知識の検証が可能となる．

第 4 章 授業内発想支援システムでは，授業内重要単語をランキングとしてリスト化するなど，人間に理解しやすい解釈方法を提供しており，この部分の整備を行っているといえる．また，第 5 章 CM 字コンテ好感度予測システムについても，出力された予測結果をベースに入力を調整できるシステムにしていることから，この部分に該当する．

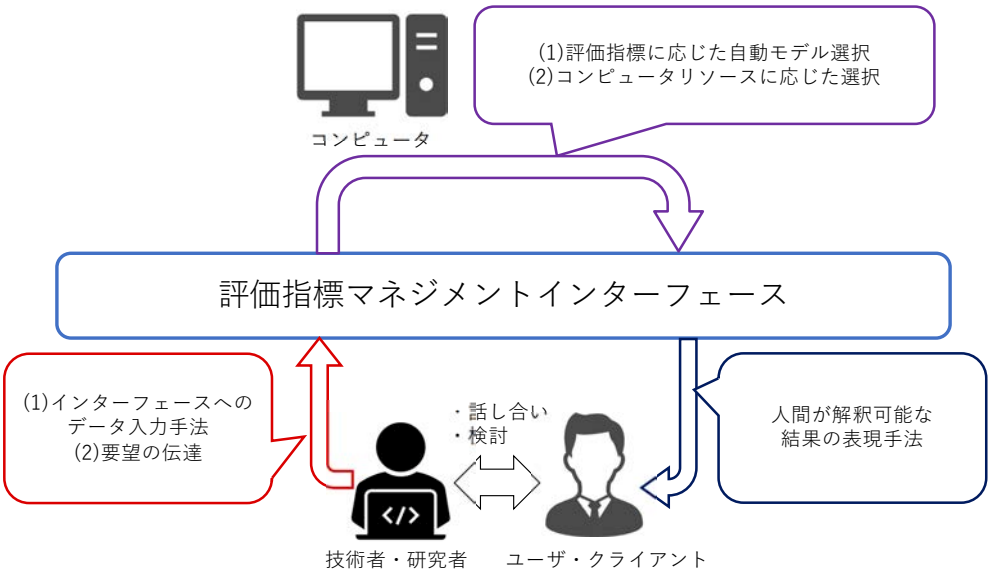


図 2.4: インターフェースと接触する人間が 2 箇所以上になる例

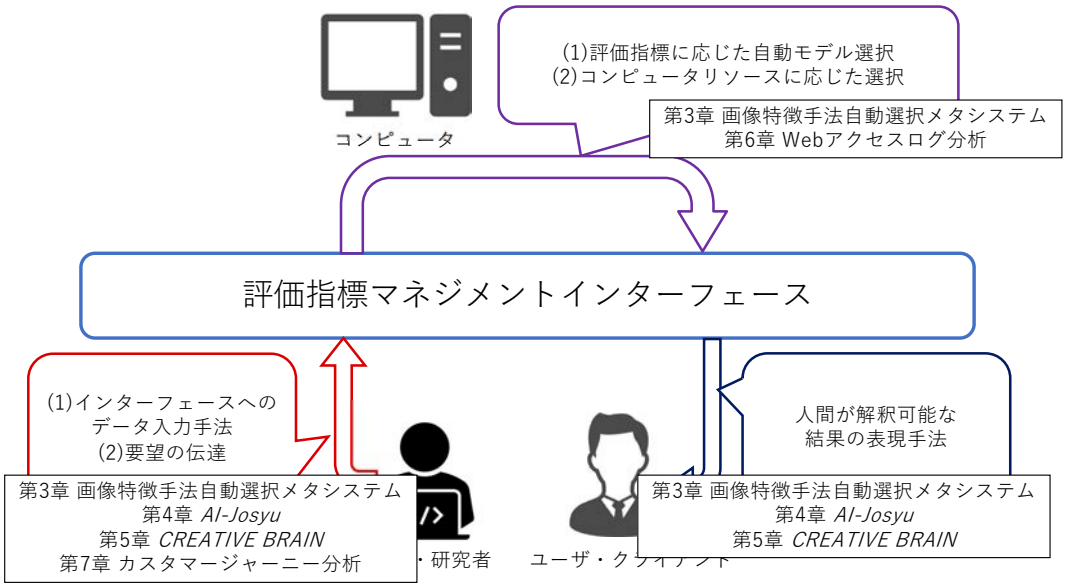


図 2.5: インターフェースの要素と本論文の内容の対応

第3章　画像特徴手法自動選択メタシステム

3.1　研究の背景

モバイルデバイスの普及によって，誰でも写真を撮影し，インターネットを通して共有できるようになった．これによって，インターネット上には急激な速度で画像データが蓄積されている．この膨大な画像データを，巨大な画像データベースとして捉え，利用することが重要になってきていると言われている [8]．この画像データベース上に存在する画像に対し，言葉として表現されたテキスト形式のメタデータを付与することによって，意味的画像データベースとして利用することが可能となる [9]．このメタデータ付与を人手によって行うことはデータの巨大さから実質的に不可能であり，その問題を解決するためには，コンピュータによって自動的に付与されるシステムが必要となる．このような自動でのタグ付けが実現すれば，コンピュータ上で画像と言葉を結びつけることができる．このメタデータ付与に応用可能な手法のひとつが画像分類である．画像の内容によってあらかじめ設定されたラベルに画像を振り分けることができれば，画像に自動的にメタデータを付与することができる．

現在，画像分類に関しては機械学習による精度向上がめざましく，特に畳み込み層を持つ深層学習 (Convolutional Neural Network, CNN) による画像分類は，2015 年には人間による分類を超えたとも報告されている [10]．しかし，機械学習による画像分類の手法が発展するまでには，様々な画像の専門家が提案してきた画像特徴量を用いて画像分類を行う手法が広く使われていた．画像特徴量とは，対象となる画像の特徴を単一または複数のベクトルで表現したものである．さらに，この画像特徴量はその特徴量を抽出する領域に応じて，画像全域から一つの特徴量を抽出する大域特徴量と，あらかじめ画像内に特徴点 (特徴領域) を複数定めておき，それぞれの特徴点から特徴量を抽出する局所特徴量に大きく分けることができる．後者の手法は画像の一部が変化してもそれ以外の部分に関しては局所特徴量が変化しないという利点から，画像同士のマッチングに広く利用されていた．局所特徴量は SIFT [11], SURF [12], ORB [13] など様々な手法が提案され，画像分類に用いることができる局所特徴量は多岐にわたっている．これらの局所特徴量は使用方法や仕組みなどは提示されているが，具体的にどのような画像，あるいはどのような目的に対して有効なのかは示されていない．そのため，ユーザは実際に局所特徴量をシステムに実装し，実装した局所特徴量が有効であるかを検証する必要がある．

本章では，画像分類手法統合のための画像特徴量の自動選択方式を示す．本方式は，多種多様な画像分類の対象に対し，画像分類の手法を統合し適用することで高精度な分類を行うことを目的とする．本方式によって構築されるシステムは，分類したい画像セットを入力とし，分類に有効な画像特徴量を出力とするシステムである．ここでの入力は「犬の写真セッ

トと犬のイラストセット」「パンダの写真セットとコアラの写真セット」などの異なる種類の画像セットとする．このようなシステムがあれば，ユーザは画像特徴量の性質を詳しく知ることなく画像特徴量を選択し，より有効な画像分類システムを構築することが可能となる．

また提案方式は，単一の画像特徴量だけではなく，必要に応じて複数の画像特徴量を組み合わせて，その組み合わせをユーザに提示する．画像特徴量を組み合わせることにより，単一の画像特徴量を用いるよりも，ユーザの様々な入力画像に対して有効に分類が行える可能性がある．この画像特徴量の組み合わせには，同一のキーポイントから出力された局所特徴量を連結して類似度を計量する方法を提案する．

しかし，この連結したベクトルを使って画像特徴量の類似度を計量する際に，局所特徴量の出力形式の差によって，次の 2 つの問題が発生する．1 つ目の問題は，局所特徴量によって，出力されるベクトルの要素数が異なることである．要素数が異なる局所特徴量を連結すると，ベクトルの要素数が多い特徴量ほど，連結後のベクトルを用いたキーポイントのマッチングに大きな影響を与えることが予想される．また 2 つ目の問題は，局所特徴量によって，出力される局所特徴量の要素が取りうる値の範囲が異なることである．画像特徴量には，実数値を要素とするもの，正の値しか取らないもの，0 か 1 の 2 通りの数字を要素とするものなど様々な手法が存在するため，要素が取りうる値の範囲が大きい局所特徴量ほど，キーポイントのマッチングに大きく影響することが予想される．

これらの問題を解決し，画像特徴量の組み合わせを実現するため，Bag-of-Keypoints [14] を利用して局所特徴量を統一的に扱う手法を提案する．Bag-of-Keypoints は画像を画像特徴量の集合とみなす考え方に基いて考案された手法であり，画像分類に応用されている．これによって局所特徴量の形式の差を吸収して連結可能な局所特徴量を構成することが実現できると考えられる．

これらに加え，この提案方式における画像特徴量の組み合わせの評価について，全探索を行わずに探索を行う方式を提案する．キーポイント検出器と局所特徴量の組み合わせを全探索して最適な手法の組み合わせを提示する場合，実装されているキーポイント検出器や局所特徴量の数に応じて探索数が指数爆発を起こし，探索が終わらない可能性がある．これを解決するため，組み合わせるキーポイント検出器や局所特徴量の数に制限を設定する方式と，組み合わせに含めるキーポイント検出器や局所特徴量のフィルタリングを行う方式を提案する．これらの方式を実装して組み合わせの枝刈りを行うことで，探索数が指数爆発することなく組み合わせの評価が行える．

本章では提案方式のシステムを実装し，自動選択方式による画像特徴量の画像分類の性能を検証する．この実験により，画像特徴量の自動選択が可能であることを示す．また，組み合わせの枝刈りによって全探索を行わない場合にも，提案方式である自動選択方式が有効であることを実験により検証する．

表 3.1: 各画像特徴量が行う処理と出力形式

手法名	キーポイント検出	特徴量記述	出力形式
SIFT	DoG	○	128 次元ベクトル
SURF	Hessian	○	64 次元ベクトル
HARRIS	HARRIS	-	-
ORB	FAST	○	256 次元ベクトル (要素が 0 または 1)

3.2 関連研究

本節では，関連研究として画像特徴量とその抽出手法，特徴選択，および Bag-of-Keypoints について述べる．

3.2.1 画像特徴量

画像特徴量とは，画像の画素値から得られる，画像の特徴を表す値である．

画像特徴量を抽出する過程は，キーポイント検出と特徴量記述の 2 つの処理に分割できる．また画像特徴量を求める手法には，2 つの処理をどちらとも行う手法と，どちらか 1 つのみ行う手法がある．本章の実験で用いている画像特徴量について，表 3.1 に示す．

キーポイント検出

キーポイント検出とは，画像の特徴が表れていると考えられる点を画像内から検出することを指す．使用するキーポイント検出の手法に応じて，得られるキーポイントの場所や数が異なる．以下にキーポイント検出の手法である DoG,Hessian 行列の近似計算,HARRIS コーナー検出器,FAST について述べる．

- DoG

  DoG(Difference of Gaussian) [11] は，画像内の輝度勾配を元にキーポイントを検出する手法である．異なるスケールでの平滑化画像を比較してエッジ検出を行い，特徴点の候補となる点を得る．その候補点のうち，画像のノイズによって発生した点や周辺画素との輝度差が少ない点を排除する．画像の背景に当たる部分のキーポイントが少なくなる傾向がある．

- Hessian 行列の近似計算によるキーポイント検出

  Hessian 行列の近似計算 [12] は，SURF のアルゴリズムで採用されている手法である．DoG によるキーポイント検出よりも高速とされている．積分画像と矩形フィルタを用いてキーポイントの検出を行う．

- HARRIS コーナー検出器

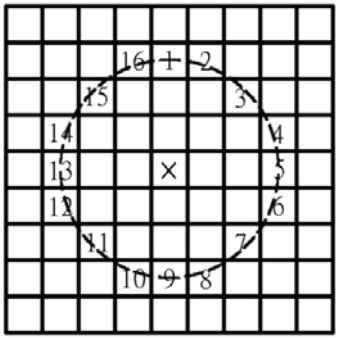


図 3.1: FAST の比較対象 16 点

HARRIS コーナー検出器 [15] は C.Harris によって提案されたコーナー検出器である．画像の縦と横に画素の輝度値を加えた 3 次元の値を曲面とみなし，Harris が提案した作用素によって曲面の曲率を求め，曲率の極値が大きい点をキーポイントとする手法である．

- FAST

  FAST(Features from Accelerated Segment Test) [16] は E.Rosten らによって提案されたコーナー検出器である．画像特徴量を映像から検出するために提案されたもので，処理が速いとされている．注目している画素から半径 3 の距離にある 16 点の画素 (図 3.1) と輝度を比較し，輝度勾配があるしきい値よりも大きい点が 9 つ以上ある点を特徴点の候補とする手法である．

特徴量記述

特徴量記述とは，画像を表す特徴を値として抽出することを指し，画像全体から特徴量を抽出する大域特徴量と，キーポイント検出で得られたキーポイントから特徴量を抽出する局所特徴量に分けることができる．抽出する特徴量には，キーポイント周辺の輝度勾配や色情報などを元に構成したベクトルを利用する．使用する特徴量記述が異なれば，出力されるベクトルの次元数や，特徴量同士での類似度の計量方法が異なる．

以下に局所特徴量の特徴量記述である SIFT,SURF,ORB について述べる．

- SIFT

  SIFT で得られる特徴量は，画像のスケールや回転に対して不変で，アフィン変換や照明変動に対し頑強であるとされている．キーポイント 1 つに対し，128 次元のベクトルの特徴量が 1 つ得られる．

  この手法では，注目している特徴点の周辺の輝度の勾配から，その特徴点が持つオリエンテーションを決定する．特徴点の周辺領域を図 3.2 のように 16 の領域に分割しし，各領域で 8 方向への輝度の勾配を求める．その輝度勾配を用いて $16 \times 8 = 128$ 次元のベ

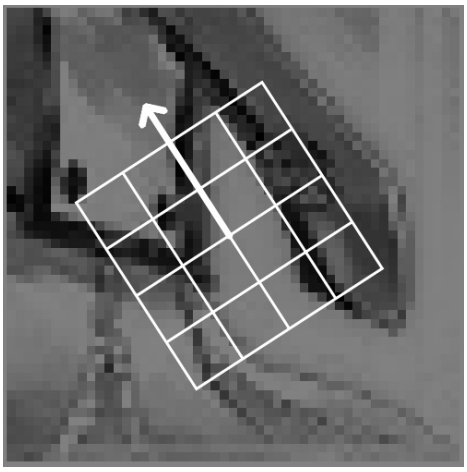


図 3.2: SIFT のオリエンテーションと 16 領域

クトルとし，特徴量とする．この際，特徴量を計算する際の特徴点の周辺領域をオリエンテーションに合わせて回転することにより，画像の回転に不変な特徴量を得ることができる．また，特徴量をそのベクトルの各値の総和で正規化するため，画像の照明条件の影響を小さくすることができる．

- SURF

  SURF は，SIFT の特徴である画像のスケールや回転に不変という性質を持ちつつ，比較的計算速度の速い手法とされている．キーポイント 1 つに対し，64 次元のベクトルの特徴量が 1 つ得られる．

  この手法では，注目している特徴点の水平方向と垂直方向に Haar Wavelet フィルタ [17] を適用し，その特徴点が持つオリエンテーションの方向を決定する．特徴点の周辺領域をオリエンテーションに合わせて 16 の領域に分割し，各領域に再度 Haar Wavelet フィルタを適用する．Haar Wavelet フィルタを用いて，各領域における水平方向・垂直方向の輝度勾配とその絶対値，合わせて 4 つの値を求める．16 の各領域で得られた 4 つの値を 1 つのベクトルとし，64 次元の特徴量を得る．

- ORB

  ORB の特徴量記述には，BRIEF [18] と呼ばれる特徴量記述を基にした手法が用いられている．キーポイント 1 つに対し，256 次元で，要素が全て 0 または 1 である特徴量が 1 つ得られる．

  この手法では，特徴点周辺の領域における 0 次モーメント・1 次モーメントの値から重心を算出し，特徴点と重心の位置からオリエンテーションを決定する．そのオリエンテーションに合わせて特徴点周辺の領域における任意の 2 点の組 (ペアと呼ぶ) を複数取り出す．このペアから代表となるものを 256 個選択し，ペアの 2 点の輝度の大小によって

0 または 1 のビットを割り当てたものが特徴量となる．本手法では 256 個のペアを学習によって選択しており，全ペアのビットの分散が大きいもの，他のペアとの相関が低いものを貪欲的に求めている．

3.2.2 画像分野における特徴選択

特徴選択は機械学習の分野で発展してきた手法であり，現在では様々な分野で広く用いられている．I. Guyon らは，特徴のランキングを作成する手法，特徴同士の相互情報量を用いる手法，特徴のクラスタリングなどが特徴選択の手法であると述べている [19].

画像における特徴選択の一つとして，G. Csurka らが提案した Bag-of-Keypoints [14] が挙げられる．また，局所特徴量における特徴選択として，長畑らは複数の局所特徴量を統一的に扱う手法を提案している [20]．長畑らはそれぞれの局所特徴量と画像の関連性を予備実験によって調査し，局所特徴量に重み付けを行うことによって特徴選択を行っている .

3.2.3 Bag-of-Keypoints

Bag-of-Keypoints [14] は G. Csurka らによって提案された手法で，画像を画素値の集合ではなく，画像特徴量の集合とみなす考え方に基づいて考案された手法である．1 枚の画像から複数の局所特徴量を抽出し，その局所特徴量の出現分布のヒストグラムで画像を表現するため，複数の局所特徴量を簡潔な表現に落とし込むことができる．局所特徴量の出現分布をヒストグラムとして構成するための基準となるベクトル群を Visual-words と呼ぶ．ヒストグラムの要素の数は Visual-words が持つベクトルの数に一致するため，Visual-words のベクトルの数を任意に決めることで，多様性を持つ局所特徴量を任意の次元数で表すことが可能となる．

もともと自然言語処理の分野において，文章を単語の出現分布のヒストグラムで表現する Bag-of-Words [21] という手法が提案されており，Bag-of-Keypoints はそれを元にして考案された手法である．Bag-of-Words が文章に適用した手法であるのに対し，画像に適用している Bag-of-Keypoints は，Bag-of-Visual-words や Bag-of-Features と表記されることがあるが，本稿では G. Csurka らの論文 [14] の表記に従い，Bag-of-Keypoints という表記に統一する．

また，Bag-of-Keypoints を用いて作られるヒストグラムは，厳密な意味でのヒストグラムではなく，実際には任意の数の値を並べて作られるベクトルである．本稿では G. Csurka らの論文 [14] の表記に従い，Bag-of-Keypoints を用いて構成するベクトルを “ヒストグラム” と表記する．そのため，本稿で述べる “ヒストグラム” の類似度とはベクトルにおける類似度と同一であり，ヒストグラムインターセクションなどのヒストグラムに用いられる類似度の他にも，ユークリッド距離やコサイン尺度など，ベクトルに使われる類似度も用いられることに注意されたい．

Bag-of-Keypoints は，Visual-words を作成する行程と，作成した Visual-words を用いて画像を局所特徴量の出現分布のヒストグラムで表現する行程の 2 つから成る．以下に，その詳細を述べる．

局所特徴量を用いた Visual-words の作成

Bag-of-Words では，文章に含まれる単語を単語群の出現分布にすることで文章からベクトルを構成する．この Bag-of-Words における単語群に相当するのが Visual-words であり，画像における語彙に相当する．この Visual-words は，画像から抽出される様々な局所特徴量から代表とする局所特徴量を複数選択し，それらを集めて作成される．Bag-of-Keypoints では，画像を Visual-words の出現分布のヒストグラムで表現するため，あらかじめどの局所特徴量を Visual-words にするかを決定する必要がある．Visual-words を作成する手順を以下に示す．また，その概要図を図 3.3 に示す．

1. 局所特徴量の抽出
   複数の任意の画像を学習画像として用意し，キーポイント検出と特徴量記述を行う．G. Csurka らの研究では，SIFT によるキーポイント検出と特徴量記述を採用している．

2. 局所特徴量のクラスタリング
   抽出された全ての局所特徴量を対象にクラスタリングを行う．この際に形成されるクラスタ数が Visual-words のベクトルの数となる．

3. Visual-words の出力
   各クラスタに含まれる局所特徴量の平均を求め，クラスタ中心とする．そのクラスタ中心を Visual-words として出力する．

Visual-words を要素とするヒストグラムへの変換

上記の手順で作成された Visual-words を用いて画像を出現分布のヒストグラムに変換する．複数の画像をヒストグラムに変換し，ヒストグラム同士の類似度を計量することにより，画像の類似度を計量することが可能となる．ヒストグラムへの変換の手順を以下に示す．また，その概要図を図 3.4 に示す．

1. 局所特徴量の抽出
   1 枚の任意の画像を用意し，キーポイント検出と特徴量記述を行う．これにより，1 枚の画像から使用する手法に対応した複数の局所特徴量を得る．

2. 局所特徴量のヒストグラム化
   ヒストグラムを構成するとき，投票と呼ばれる処理を行う．作成するヒストグラムは，Visual-words に含まれるベクトルと同じ数のビンを持つ．投票では，画像から抽出された任意の局所特徴量 1 つを取り出して，その局所特徴量と最近傍となるベクトルを Visual-words の中から探索し，そのベクトルに対応するビンに 1 を加える．この投票を 1 枚の画像から得られる全ての局所特徴量に対して行い，ヒストグラムを構成する．

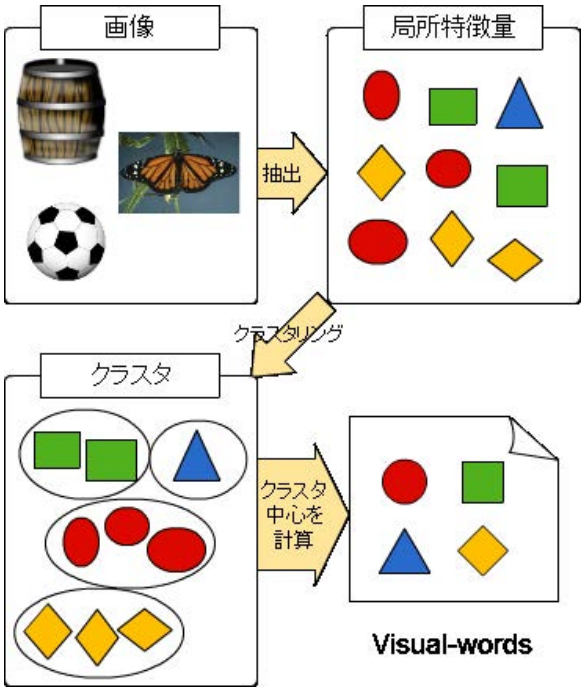


図 3.3: Visual-words の作成

3. ヒストグラムの正規化
   1 枚の画像から得られる局所特徴量の数はキーポイントの数に依存し，画像によりその数は異なる．そのため，ヒストグラムの各要素を，その画像から得られた局所特徴量の数で割って正規化する．

3.3 画像特徴量の自動選択方式

本節では提案方式の画像特徴量の自動選択方式のモデルと処理について述べる．

3.3.1 画像特徴量の自動選択方式の概要

提案方式のモデルを 3.5 に示す．ユーザは画像分類を行いたい画像セットから学習用画像セットを複数選択し，その学習用画像セットに対して手動で分類を行う．その手動で分類を行った学習用画像セットを，画像分類の正解セットとしてシステムに入力する．システムはその画像セットに応じて画像特徴量を選択し，必要に応じて画像特徴量を組み合わせ，学習用画像セットを最も良く画像分類できる画像特徴量の組み合わせをユーザに提示する．この方式によって提示された画像特徴量の組み合わせをユーザが準備した画像セットに適用する

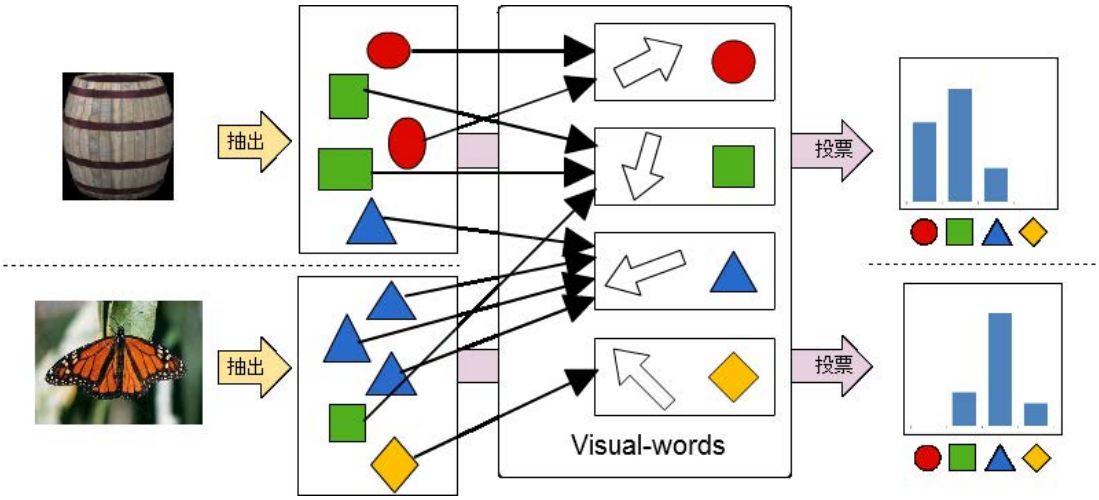


図 3.4: 出現分布のヒストグラムへの変換

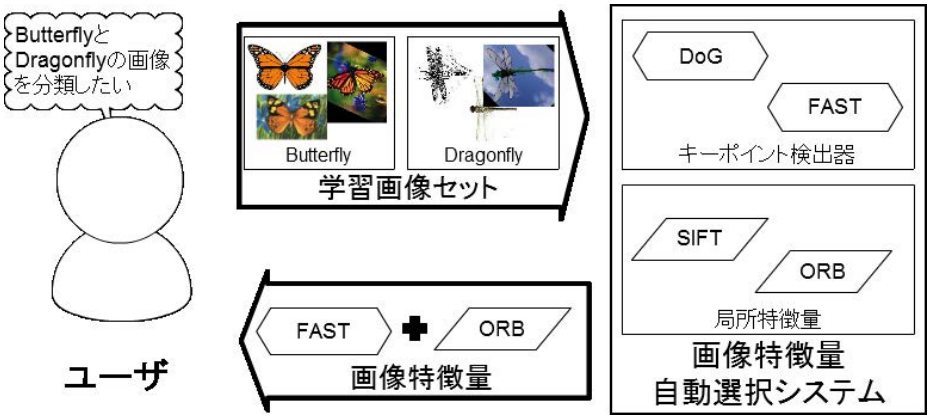


図 3.5: 画像特徴量の自動選択システムのモデル

ことで，画像セットを手動で分類した学習用画像セットと同じように高精度で分類されることが期待できる．

3.3.2 提案方式における Visual-words の作成

画像から抽出した複数の局所特徴量をヒストグラムに変換する際，Bag-of-Keypoints では Visual-words を用いる．ヒストグラムに変換することによって，特徴量記述によって異なっていた局所特徴量の要素数を統一することができる．この Visual-words の作成は，システムに実装する特徴量記述ごとに行う．この Visual-words の作成には既存の手法を用いるために，詳細な手順は省き，大まかな手順だけを述べる．詳細な手順は Bag-of-Keypoints の論文を参照されたい [14].

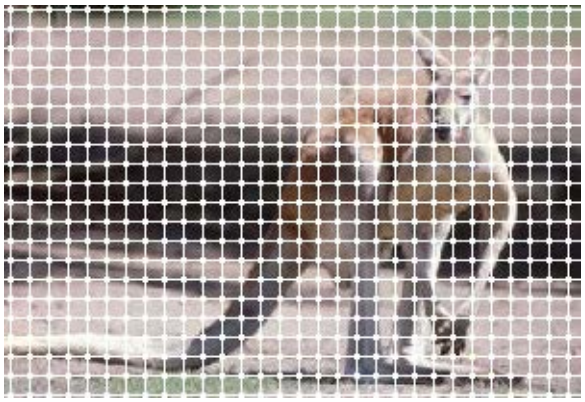


図 3.6: Grid-sampling

1. キーポイント検出
   キーポイント検出には Grid-sampling（図 3.6，Dense-sampling とも）を用いる．Grid-sampling は，一定間隔で配置された格子点をキーポイントとするサンプリング手法であり，Visual-words の作成においては，DoG などのキーポイント検出器を使用するよりも有利であるという研究結果がある [22].

2. 特徴量記述
   Grid-sampling で得たキーポイントを用いて，特徴量記述を行う．

3. 局所特徴量のクラスタリング
   局所特徴量のクラスタリングは，特徴量記述ごとに行う．抽出した局所特徴量を k-means++法 [23] でクラスタリングし，各クラスタ中心を Visual-words として出力する．このクラスタリングにおけるクラスタ数が，Bag-of-Keypoints を用いて構成したヒストグラムの要素数となる．そのため，特徴量記述に関わらずクラスタ数を一定にしてクラスタリングを行うことで，特徴量記述に応じて異なるベクトルの次元数を統一することができる．

4. 特徴量記述に対応した Visual-words の作成
   特徴量記述の手法ごとにこれらの行程を行い，特徴量記述ごとに 1 つの Visual-words を作成する．例えば，提案システムに SIFT,SURF,ORB の 3 種類を実装したい場合には，SIFT 用の Visual-words，SURF 用の Visual-words，ORB 用の Visual-words の 3 種類を作成する．

3.3.3 キーポイント検出器・局所特徴量の組み合わせ探索の枝刈り方式

キーポイント検出器と局所特徴量の組み合わせを評価する際，キーポイント検出器や局所特徴量の数に応じて組み合わせ数が爆発的に増える．システムに組み込まれているキーポイ

ント検出器の数を α，局所特徴の抽出手法の数を β としたとき，手法の組み合わせの数は O $(2^{\alpha+\beta})$ となり，探索が終わらない可能性がある．

そのため，画像特徴量の自動選択方式が組み合わせの全探索を行わずに画像特徴量の組み合わせの提示ができるように，以下の 2 つの方法を提案している．

組み合わせるキーポイント検出器・局所特徴量の数の制限

キーポイント検出と局所特徴の抽出手法を組み合わせる際，その組み合わせに用いる手法の数に上限を設定する．最大数 n を設定することにより，キーポイント検出器が α 種，局所特徴量が β 種実装されているシステムならば組み合わせの数を O $((\alpha+\beta)^n)$ に抑えることができる．

これはキーポイント検出に対して特に有効な手段であるといえる．なぜならば，キーポイントが多いほど有効であるとは必ずしも言えないからである．例えば，G. Lowe らが提案している DoG [11] では一度検出したキーポイントについて，キーポイントの周りの輝度勾配が少ないなどの不適と思われるキーポイントを除外する処理がある．

また，組み合わせる手法の数が多いほど組み合わせの評価に必要な時間が伸びる傾向があるため，組み合わせる数が多いものを積極的に除外するほうが処理時間を短縮できると考えられる．

キーポイント検出器・局所特徴量のランキングによるフィルタリング

組み合わせて評価を行う前に，キーポイント検出，局所特徴の抽出手法をそれぞれ単独で実行して評価関数を求め，その結果をランキングにする．そのうち下位にあるものは組み合わせから除外し，評価を行わないようにする．

ランキングから m 種取り出し，それら以外を組み合わせから除外しつつ評価を行っていくと，評価する組み合わせの数を O (2^m) に抑えることができる．

3.4 提案方式による画像特徴量の自動選択システム

本章では，第 3 章の提案方式を元に構成するシステムの概要を示す．

提案方式は，分類したい画像セットから学習用画像セットを複数選択してを入力とし，その学習用画像セットを最も良く分類できる画像特徴量を自動選択する方式である．この方式で選択された画像特徴量は，学習用画像セットと同じように画像セットを高精度で分類することが期待できる．

本方式において，異なる局所特徴量を組み合わせることを実現するため，Bag-of-Keypoints を使用する．これにより，Visual-words を用いたヒストグラムに変換することで，局所特徴量の要素数の差を吸収することに加え，局所特徴量ごとに異なる要素の構成方法を統一することができ，連結が可能な局所特徴量を作成できる．提案方式の処理の流れを図 3.7 に示す．以下に，画像特徴量の自動選択方式の処理の詳細を述べる．

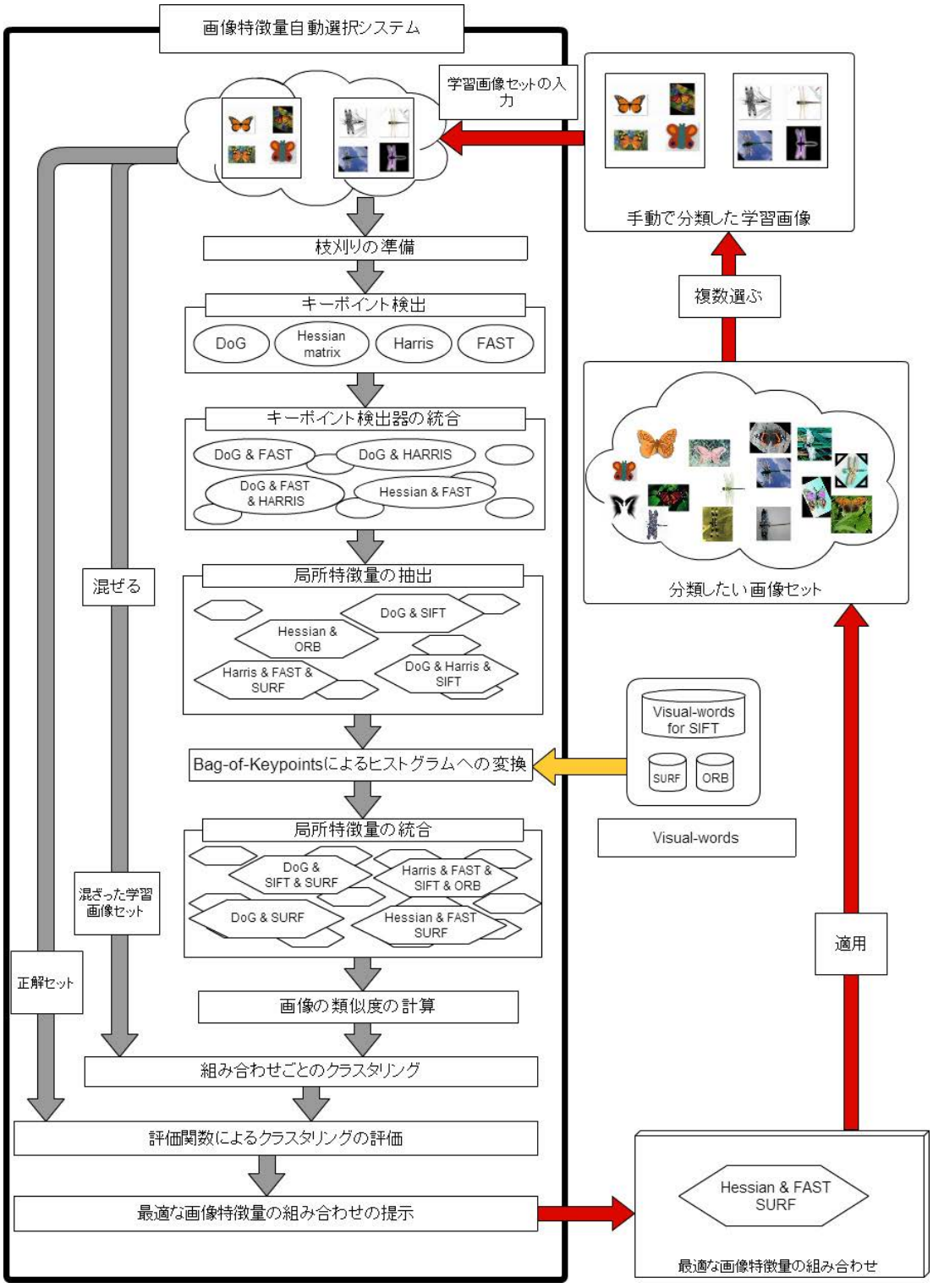


図 3.7: 画像特徴量の自動選択の処理の流れ

3.4.1 学習用画像セットの入力

本方式を利用するユーザは，分類したい画像セットから複数の学習用画像セットを選択し，それらを手動で分類する．この学習用画像セットは後述する画像クラスタリングにおいて，クラスタリング結果と比較する正解セットとして利用する．この正解セットにより近い分類結果を示す画像特徴量を自動選択し，ユーザに提示することが本方式の目的となる．

3.4.2 枝刈りのための準備

本方式で提案した 2 種類の枝刈りを行う．

組み合わせるキーポイント検出器・局所特徴量の数の制限

組み合わせるキーポイント検出器・局所特徴量の最大数 n を設定する．キーポイント検出器が α 種，局所特徴量が β 種実装されているシステムならば組み合わせの数を $\mathrm{O}\left((\alpha+\beta)^n\right)$ に抑えることができる．

キーポイント検出器・局所特徴量のランキングによるフィルタリング

ランキング上位から取り出すキーポイント検出器・局所特徴量の数 m を設定する．ランキングから m 種取り出し，それら以外を組み合わせから除外しつつ評価を行っていくと，評価する組み合わせの数を $\mathrm{O}\left(2^m\right)$ に抑えることができる．

3.4.3 キーポイント検出

入力された画像に対し，システムに実装されているそれぞれのキーポイント検出を行う．単一のキーポイント検出器で得られた全てのキーポイントを 1 つのキーポイント集合とし，キーポイント検出器の数だけキーポイント集合を得る．キーポイント検出器によって検出されるキーポイントの数や位置が異なるため，キーポイント検出器の処理の内容を反映した様々なキーポイント集合が得られる．

3.4.4 キーポイント検出器の組み合わせ

2 つ以上のキーポイント集合を統合して 1 つのキーポイント集合とし，統合したキーポイント集合をキーポイント検出器の組み合わせで検出したものと考える．これは，同一の画像に対して複数のキーポイント検出器で得られた結果と等しくなる．そのため，各キーポイント検出器で得られたキーポイント集合の和集合を求めることによって組み合わせを実現する（図 3.8）．これによって，複数のキーポイント検出器の性質を取り入れたキーポイント集合が得られる．ここで，キーポイント検出器が α だけ実装されているシステムならば，キーポ

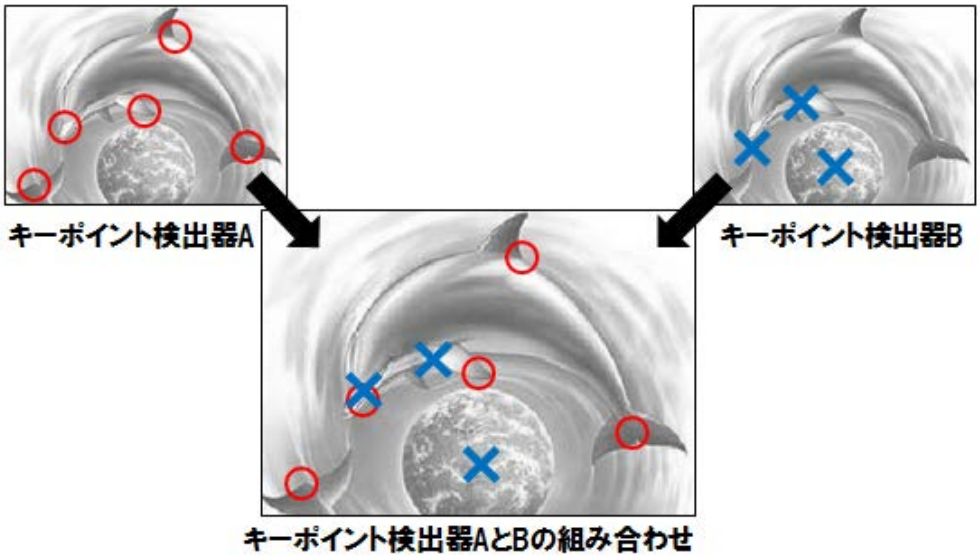


図 3.8: キーポイント検出器の組み合わせの実現方法

イント検出を 1 つも使わない場合を除き，最大で $(2^{\alpha}-1)$ 通りのキーポイント検出器の組み合わせが得られる．

3.4.5 特徴量記述

キーポイント検出で得られたそれぞれのキーポイント集合に対し，システムに実装されているそれぞれの特徴量記述を行う．1 つのキーポイント集合に対しキーポイントの数と同じだけの局所特徴量が抽出されるので，キーポイント集合と特徴量記述の組み合わせを変えながら，組み合わせごとに結果を保存する．ここで組み合わせを変化させることにより，キーポイント検出器と特徴量記述の様々な組み合わせを実現している．

3.4.6 Visual-words を要素とするベクトル表現への変換

特徴量記述によって得られた局所特徴量を用いて，3.2.3 で述べた投票の操作を行い，画像を Visual-words を要素としたヒストグラムに変換する．変換後のヒストグラムの要素数は Visual-words のベクトル数に依存するため，様々な次元数の局所特徴量が同じ次元数に統一される．さらに，ヒストグラムの各要素は投票によって求めるため，局所特徴量の抽出手法に関わらず全ての要素が 0〜1 の値に統一される．このヒストグラムへの変換を特徴量記述ごとに行い，画像 1 枚に対して特徴量記述の組み合わせの数と同じだけのヒストグラムを構成する．ここまでの処理により，キーポイント検出器が α，特徴量記述が β だけ実装されているシステムならば，1 枚の画像から 最大で $(2^{\alpha}-1)\cdot\beta$ 通りのヒストグラムが得られる．

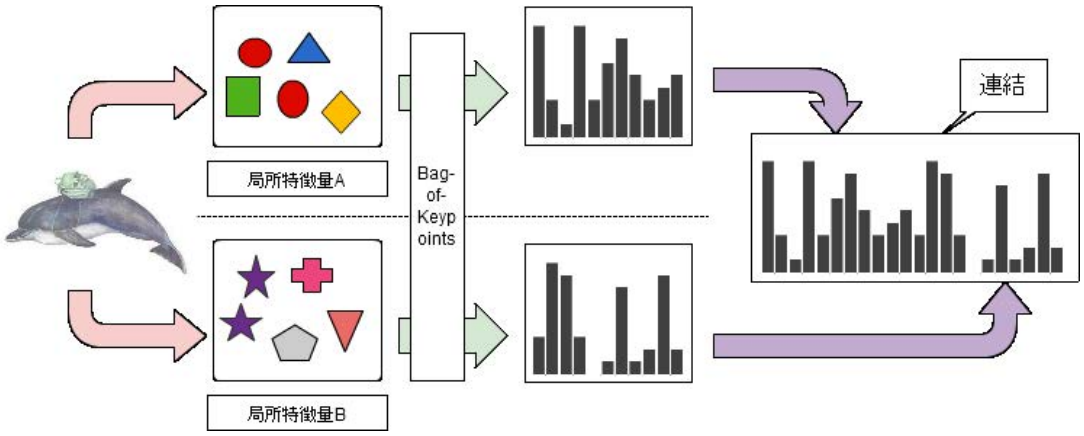


図 3.9: 特徴量記述の組み合わせの実現方法

3.4.7 特徴量記述の組み合わせ

3.4.6 で得られたヒストグラムを，組み合わせる特徴量記述に対応するように連結したものを，2 つ以上の特徴量記述を組み合わせた場合のヒストグラムとする（図 3.9）．ヒストグラムの要素数や，その要素の値が取りうる範囲が統一化されているため，特定の局所特徴量の要素が大きく突出することなく連結することが可能となる．この連結したヒストグラムを用いて，画像の類似度計量を行う．特徴量記述が β だけ実装されているシステムならば，変換したベクトルを 1 つも使わない場合を除き，最大で $(2^{\beta}-1)$ 通りの連結方法が存在する．評価実験に用いた提案方式によるシステムでは，SIFT や SURF などの代表的な画像特徴量に倣い，各画像を変換した 2 つのヒストグラムのユークリッド距離を 2 枚の画像の類似度をとしている．

3.4.8 類似度による画像クラスタリング

キーポイント検出器が α，特徴量記述が β だけ実装されているシステムならば，ここまでの処理によって，$(2^{\alpha}-1)\cdot(2^{\beta}-1)$ の組み合わせを実現することができる．3.4.7 で求められた類似度を用いて，入力された学習用画像セットを画像特徴量の組み合わせごとにクラスタリングする．このときのクラスタリングのクラスタ数は入力として与えた正解セットのグループ数とする．本稿の実験システムでは，クラスタリング結果のクラスタ数を指定できる k-means 法を用いている．

3.4.9 クラスタリング結果の評価

クラスタリング結果に対し，評価関数を用いて評価を行う．本稿の実験システムでは以下の 2 つの評価関数を採用している．

1. Purity

   Purity はクラスタリング結果の各クラスタごとの多数派が占める割合である．以下の式で定義される．

$$Pu(C_i) = \frac{1}{n_i} \max n_i^j.$$

$$purity = \sum_{i=1}^{k} \frac{n_i}{n} Pu(C_i).$$

   ここで C_i は i 番目のクラスタ，n_i は i 番目のクラスタのデータ数，n_i^j はクラスタ i のうち学習用画像セット j のデータ数，k は全クラスタ数，n は全データ数である．$purity$ が大きいほど良いクラスタリング結果といえる．

2. Entropy

   Entropy はクラスタリング結果の各クラスタの乱雑さを表す．以下の式で定義される．

$$E(C_i) = -\frac{1}{\log q} \sum_{j=1}^{q} \frac{n_i^j}{n_i} \log \frac{n_i^j}{n_i}.$$

$$entropy = \sum_{i=1}^{k} \frac{n_i}{n} E(C_i).$$

   ここで q は学習画像セットのグループ数である．ただし提案方式では必ず $q = k$ となる．他の変数は Purity と同様に定義される．$entropy$ が小さいほど良いクラスタリング結果といえる．

3.4.10 画像特徴量の組み合わせの選択

評価関数それぞれについて，最も良いクラスタリングを行った画像特徴量の組み合わせをユーザに提示する．評価関数が同じ値を示し，評価関数のみでは組み合わせの優劣がつけられない場合には，キーポイントの検出や局所特徴量の抽出にかかる処理時間が短いものを提示する．これは，画像特徴量の利用例の 1 つに動画像からのリアルタイムな局所特徴量の抽出があることや，FAST [16]，SURF [12], BRIEF [18] など精度よりも速度を重視した画像特徴量が提案されており，画像特徴量の抽出に速度を求める背景があるからである．

3.5 実験

3.5.1 画像データセット

提案方式を実装した提案システムの検証実験に際し，Caltech 101 と呼ばれる画像データセットを用いる．このデータセットはカリフォルニア大学の有志によって集められた画像データであり，画像分類の実験において広く利用されている [24].

Caltech 101 は 101 個の画像カテゴリに加え，背景画像のみを集めたカテゴリ 1 個の計 102 カテゴリ，全 9145 枚の画像で構成されている．1 カテゴリの画像は少なくとも 30 枚で構成されていることが保証され，また，どの画像も 300px × 200px 程度の解像度に調整されている．

本実験では 102 カテゴリのうち,「BACKGROUND_Google」と顔認識に用いられる「Faces」,「Faces_easy」の 3 つのカテゴリを除いた 99 カテゴリの画像を実験に使用する (図 3.10).「BACKGROUND_Google」を除外したのは，このカテゴリに含まれる画像の内容に統一性のなく，画像分類という目的から外れているという理由からである．また,「Faces_easy」に含まれる画像は,「Faces」の各顔画像から背景を取り除いたものであり,「Faces」と「Faces_easy」の内容はほぼ同じものであるため，今回の実験システムの入力としては不適と考えて入力から除外した．

3.5.2 実験条件

ここでは提案システムに設定する条件について述べる．

Visual-words の作成の条件

Visual-words の作成について，以下のように各条件を設定する．Visual-words のベクトルの数に関して予備実験を行い，極端に少ない数でなければシステムを構築できることを確かめた．(予備実験については末尾の付録に記載．) その上で，各条件を変えた場合の実験結果の変化を確認するため，Visual-word の作成にかかる時間が長時間にならないように，最終的な Visual-words のベクトル数を 1000 とした．その値に基づき，各条件を決定した．

- k-means++法におけるクラスタ数を 1000 とする．これにより，Visual-words のベクトル数は 1000 になる．
- Grid-sampling の間隔は x,y 方向ともに 6 ピクセルとする．どの画像も 300px × 200px 程度の解像度に調整されているため，1 枚の画像から 1670 個前後のキーポイントが得られる．
- クラスタ数 1000 よりも十分に大きく，クラスタを構成するのに不都合がない程度の数のキーポイントを得るため，各カテゴリからランダムに 1 枚ずつ抜き出した 99 枚の画像を用いる．1 枚の画像につき 1670 個前後のキーポイントが得られるので，99 枚の画像で 165000 個程度のキーポイントとなり，クラスタ数 1000 に比べて十分に大きい数が得られる．
- 使用する局所特徴量は SIFT,SURF の 2 種類とする．提案方式では実装する画像特徴量の組み合わせ全てを評価するため，実装する特徴量記述が多いと現実的な時間内に処理が終わらない可能性がある．それを回避するため，代表的な特徴量記述 2 種類のみを実装する．

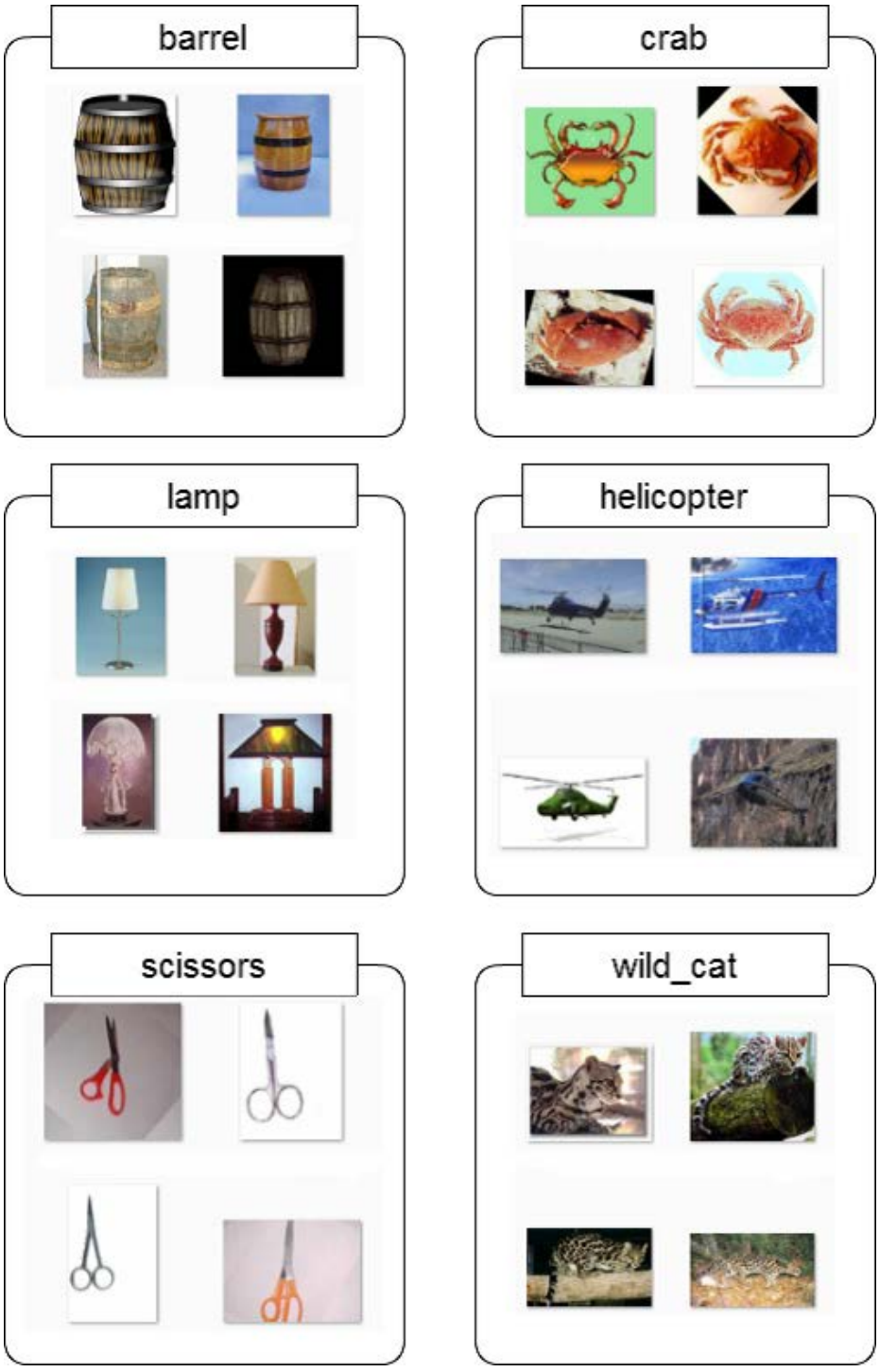


図 3.10: 学習画像の例

画像特徴量の自動選択システムの条件

画像特徴量の自動選択システムについて，以下のように各条件を設定する．

- 入力する画像セットは 2 種類とする．これは，システムをなるべく小さい構成とするためである．また，2 種類の画像を分類可能であれば，本方式を多段にすることによって，3 種類以上の分類も可能である．
- 使用するキーポイント検出器は DoG, Hessian, HARRIS, FAST の 4 種類とする．組み合わせの数が膨大になることを防ぐため，今回の実験では代表的な 4 種類のみを実装する．
- 使用する局所特徴量は SIFT, SURF の 2 種類とする．ここで利用できる局所特徴量は Visual-words を作成してあるものに限られるため，Visual-words の作成の際と同じ理由により 2 種類のみの実装とする．

キーポイント検出器が 4 種類，特徴量記述が 2 種類実装されている提案システムでは，評価するキーポイント検出・特徴量記述の組み合わせは 45 通りとなる．

3.5.3 実験

提案システムの有効性を確かめるため，検証実験を行う．

実験 1 画像特徴量の自動選択の検証

実験により，提案方式が画像特徴量を自動選択し，ユーザに提示できることを検証する．また，入力する学習画像セットに応じて，提示する画像特徴量が変化することを検証する．

準備した 99 カテゴリから 2 種類のカテゴリを選び，各カテゴリから 50 枚ずつ画像を取り出して入力とした．ここで選んだカテゴリは以下の 3 通りとする．この組み合わせは，準備したカテゴリから無作為に選んだものである．

- 「accordion」と「airplanes」
- 「kangaroo」と「leopards」
- 「chandelier」と「dolphin」

アコーディオンと飛行機の実験に使用している画像の一例を図 3.11 に示す．

提案方式では画像特徴量の組み合わせを出力としているが，今回の実験では全ての組み合わせに対する評価関数の値を出力とするようにし，評価関数の値を比較できるようにしている．

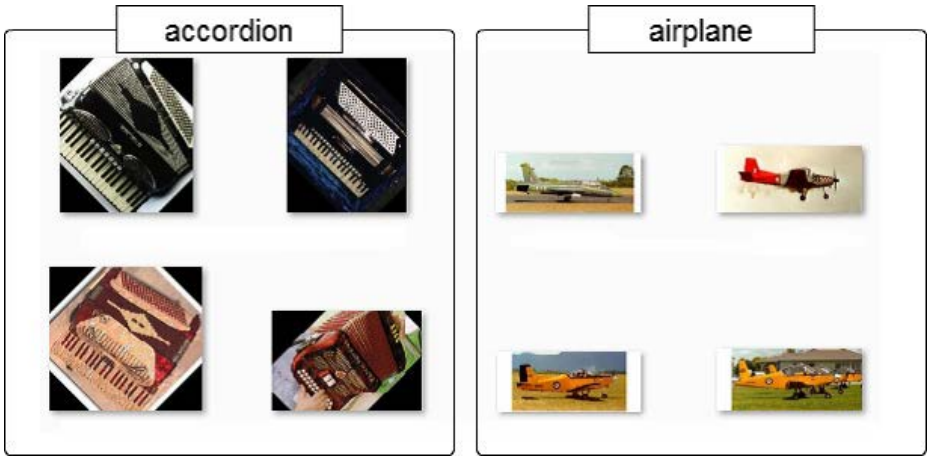


図 3.11: 実験に使用したアコーディオンと飛行機の画像の一例

実験 2 k-分割交差検証による分類精度の検証

「accordion」と「airplanes」のカテゴリについて，提示される画像特徴量の組み合わせが高精度で分類を行えることを示すために k-分割交差検証を行う．

本実験では学習用画像セットを提案システムに入力して提示された画像特徴量を用いて，検証用画像セットを分類する．ここでの分割数は 5 とし，「accordion」と「airplanes」から各 40 枚ずつ計 80 枚を取り出したものを学習用画像セット，残りの 10 枚ずつ計 20 枚を検証用画像セットとする．

まず，すべての学習用画像セットを Bag-of-Keypoints を用いてヒストグラムへ変換する．カテゴリごとの 40 枚についてヒストグラムの平均を求め，そのカテゴリを表すヒストグラムとする．その後，検証用画像セットについても Bag-of-Keypoints を用いてヒストグラムへ変換する．検証用画像セットを 1 枚ずつ取り出し，カテゴリを表すヒストグラムと比較し，ユークリッド距離が近いカテゴリに分類する．この分類を検証用画像セット 20 枚すべてについて行う．

このときの正解率を以下の式で定義する．

$$(\text{正解率}) = \frac{(\text{正しく分類された画像の枚数})}{(\text{検証用画像セットの枚数})}.$$

システムが提示する画像特徴量の組み合わせを用いて行う分類に加え，比較として，G. Csurka らが行っている DoG と SIFT を用いる場合の分類精度についても検証を行う．

実験 3 提案方式における枝刈りの有効性の検証

提案方式における枝刈りを行い，枝刈りの処理を行った場合でも画像特徴量の組み合わせが提示できることを検証する．

枝刈りに関連するパラメータは以下のように設定する．

表 3.2: 「airplane」と「accordion」を入力したときの Purity,Entropy

評価関数				Purity			Entropy		
特徴量記述 →				SIFT		SIFT	SIFT		SIFT
↓キーポイント検出					SURF	SURF		SURF	SURF
DoG				0.90	0.80	0.93	0.449	0.683	0.365
	Hessian			0.82	0.82	0.89	0.677	0.677	0.488
DoG	Hessian			0.92	0.92	0.89	0.390	0.390	0.496
		HARRIS		0.99	0.99	0.99	0.071	0.071	0.071
DoG		HARRIS		0.93	0.93	0.93	0.343	0.342	0.365
	Hessian	HARRIS		0.93	0.93	0.90	0.358	0.358	0.460
DoG	Hessian	HARRIS		0.92	0.92	0.87	0.390	0.390	0.539
			FAST	0.87	0.87	0.87	0.539	0.539	0.539
DoG			FAST	0.93	0.93	0.94	0.365	0.365	0.327
	Hessian		FAST	0.88	0.88	0.89	0.529	0.529	0.499
DoG	Hessian		FAST	0.90	0.90	0.96	0.467	0.467	0.242
		HARRIS	FAST	0.92	0.92	0.93	0.399	0.399	0.365
DoG		HARRIS	FAST	0.94	0.94	0.96	0.327	0.327	0.242
	Hessian	HARRIS	FAST	0.91	0.91	0.92	0.436	0.436	0.390
DoG	Hessian	HARRIS	FAST	0.93	0.93	0.94	0.365	0.365	0.327

1. 組み合わせる画像特徴量の最大数 n : 4

2. ランキングの上位から取り出す画像特徴量の数 m : 4

ランキングの上位から取り出す画像特徴量の数 m を設定した場合の，単一のキーポイント検出器と単一の局所特徴量を用いた場合のランキングを出力し，実験 1 における「accordion」「airplane」の組み合わせについて同様に画像特徴量の自動選択を行う．

3.5.4 実験結果

実験 1 の結果

表 3.2，表 3.3，表 3.4 にそれぞれ結果を示す．表の行はそれぞれ使用したキーポイント検出の組み合わせで分けられており，記入されているアルゴリズム名が組み合わせに含まれているものとなる．表の列は特徴量記述について，行の場合と同じように記述されている．例えば，1 行 3 列に記入されている 0.93 という値は，キーポイント検出器に DoG を使い，特徴量記述に SIFT と SURF を組み合わせたものを使った場合の Purity の値を指している．さらに，最も良い数値を下線で示している．

表 3.3: 「kangaroo」と「leopards」を入力したときの Purity

				評価関数	Purity	
				特徴量記述 →	SIFT	SIFT
↓キーポイント検出					SURF	SURF
DoG					0.63	0.63	0.64
	Hessian				0.94	0.94	0.94
DoG	Hessian				0.94	0.94	0.59
		HARRIS			0.52	0.55	0.62
DoG		HARRIS			0.64	0.65	0.58
	Hessian	HARRIS			0.95	0.95	0.65
DoG	Hessian	HARRIS			0.95	0.95	0.60
			FAST		0.82	0.79	0.82
DoG			FAST		0.75	0.75	0.75
	Hessian		FAST		0.84	0.84	0.84
DoG	Hessian		FAST		0.82	0.82	0.84
		HARRIS	FAST		0.84	0.84	0.83
DoG		HARRIS	FAST		0.75	0.75	0.75
	Hessian	HARRIS	FAST		0.95	0.95	0.86
DoG	Hessian	HARRIS	FAST		0.84	0.84	0.83

表 3.2 の「airplane」と「accordion」の結果では，キーポイント検出に HARRIS，特徴量記述に SIFT,SURF,SIFT と SURF のどの 3 種類を使った場合にも Purity が最大の 0.99 となり，この 3 種類が最も良い組み合わせとなることがわかる．今回の実験では 100 枚の画像を入力としているため，Purity が 0.99 のときには，クラスタリング結果のうち 1 枚だけが正解用として用意した学習用画像セットとは別のグループにクラスタリングされたことになる．この結果から，学習用画像セットと同じような画像を分類する際に提案された画像特徴量の組み合わせを用いることで，高精度な分類が行えることが期待される．Entropy を用いた場合でも最小の 0.071009 となり，最も良い組み合わせは同じものを示している．また，最も数値が低い結果はキーポイント検出に DoG，特徴量記述に SURF を使った場合である．最も良い場合と比べて 0.19 の差があり，正解セットとは別のグループにクラスタリングされた画像の数に 19 枚もの差があることになる．今回の実験では，どちらの評価関数を用いても組み合わせの評価は同じ順位となった．実際のシステムでは，同じ Purity の値が出た場合には処理時間が最も短いものを採用し，キーポイント検出に HARRIS，特徴量記述に SURF を使用する組み合わせを提示する．

表 3.3 の「kangaroo」と「leopards」の結果では，キーポイント検出に Hessian と HARRIS，特徴量記述に SIFT を使用した場合や，キーポイント検出に Hessian と HARRIS と FAST，特徴量記述に SIFT を使用した場合など 6 種類で最高値を得られた．また，表 3.4 の「chandelier」

表 3.4: 「chandelier」と「dolphin」を入力したときの Purity

評価関数				Purity		
特徴量記述 →				SIFT		SIFT
↓キーポイント検出					SURF	SURF
DoG				0.70	0.70	0.67
	Hessian			0.67	0.66	0.73
DoG	Hessian			0.76	0.76	0.84
		HARRIS		0.51	0.51	0.51
DoG		HARRIS		0.69	0.68	0.69
	Hessian	HARRIS		0.74	0.72	0.75
DoG	Hessian	HARRIS		0.76	0.76	0.72
			FAST	0.53	0.53	0.52
DoG			FAST	0.57	0.57	0.77
	Hessian		FAST	0.65	0.65	0.69
DoG	Hessian		FAST	0.73	0.73	0.75
		HARRIS	FAST	0.50	0.50	0.51
DoG		HARRIS	FAST	0.61	0.61	0.78
	Hessian	HARRIS	FAST	0.68	0.69	0.72
DoG	Hessian	HARRIS	FAST	0.71	0.71	0.77

と「dolphin」の結果では，キーポイント探索に DoG と Hessian，特徴量記述に SIFT と SURF を組み合わせた場合で最高値を得られた．この結果から，与える学習画像セットによって自動選択される組み合わせが異なっていることがわかる．

実験 2 の結果

表 3.5 の結果により，SIFT を使用した Bag-of-Keypoints による画像分類よりも，システムが提示する画像特徴量を使用したほうが良い結果が得られた．SIFT を用いた場合の正解率は 0.90，提示された HARRIS と SURF の組み合わせを用いた場合の正解率は 0.99 であり，今回の「accordion」と「airplanes」の画像セットの場合，9%の画像分類の正解率の向上が見られた．

実験 3 の結果

単独のキーポイント検出器・局所特徴量でのランキング結果

単一のキーポイント検出器と単一の局所特徴量を用いた場合のランキングを表 3.6 に示す．今回は上位 4 種類を組み合わせの対象とするよう設定したが，4 番目と 5 番目の特徴量が同じ値を示したため，上位 5 種類を組み合わせの対象とした．

表 3.5: k-分割交差検証による分類の正解率

	提示方式 上:キーポイント検出 下:特徴量記述	正解率	DoG SIFT	正解率
画像セット A	HARRIS SURF	1.00 (20/20)	DoG SIFT	0.90 (18/20)
画像セット B	HARRIS SURF	1.00 (20/20)	DoG SIFT	0.90 (18/20)
画像セット C	HARRIS SURF	1.00 (20/20)	DoG SIFT	0.95 (19/20)
画像セット D	HARRIS SURF	1.00 (20/20)	DoG SIFT	0.90 (18/20)
画像セット E	HARRIS SURF	0.95 (19/20)	DoG SIFT	0.85 (17/20)
平均値		0.99		0.90

表 3.6: 単一のキーポイント検出器と単一の局所特徴量を用いた場合の Purity

	キーポイント検出器	局所特徴量	Purity
	HARRIS	SIFT	0.99
	HARRIS	SURF	0.99
	DoG	SIFT	0.90
	FAST	SURF	0.87
	FAST	SIFT	0.87
↓除外	Hessian	SURF	0.82
	Hessian	SIFT	0.82
	DoG	SURF	0.80

枝刈りを行った場合の画像特徴量の自動選択の結果

表 3.7 に実験結果を示す．最も良い数値を示しているものは下線，提案方式によって評価の対象とならないものは括弧をつけている．全探索では 45 通りの組み合わせとなるが，提案方式により評価する組み合わせは 16 通りに減少している．

3.5.5 実験結果の考察

実験結果より，提案方式が画像を入力とし，画像特徴量の組み合わせを出力できることを検証した．与える学習画像セットによって，提示される画像特徴量が変化することを検証し

表 3.7: 枝刈りを行った場合の「accordion」と「airplain」の Purity

↓キーポイント検出 \ 局所特徴量 →				SIFT	SURF	SIFT SURF
DoG				0.90	0.80	(0.93)
	Hessian			0.82	0.82	(0.89)
DoG	Hessian			(0.92)	(0.92)	(0.89)
		HARRIS		0.99	0.99	0.99
DoG		HARRIS		0.93	(0.93)	(0.93)
	Hessian	HARRIS		(0.93)	(0.93)	(0.90)
DoG	Hessian	HARRIS		(0.92)	(0.92)	(0.87)
			FAST	0.87	0.87	0.87
DoG			FAST	0.93	(0.93)	(0.94)
	Hessian		FAST	(0.88)	(0.88)	(0.89)
DoG	Hessian		FAST	(0.90)	(0.90)	(0.96)
		HARRIS	FAST	0.92	0.92	0.93
DoG		HARRIS	FAST	0.94	(0.94)	(0.96)
	Hessian	HARRIS	FAST	(0.91)	(0.91)	(0.92)
DoG	Hessian	HARRIS	FAST	(0.93)	(0.93)	(0.94)

た．また，提示された画像特徴量によって画像分類を行うことにより，高精度の分類が可能であることを示した．これによって提案方式が，目標とする自動選択を行えたといえる．

どの学習画像セットを与えた場合でも，特徴量記述が SIFT のみの場合と，SURF のみの場合に出力される数値がほぼ同じであることが確認できる．SURF は SIFT の速度向上版であり，簡素化した式を使用しているために，今回の実験では大きな性能の差が出なかったでのはないかと考えられる．そのような SIFT と SURF を組み合わせた特徴量も同じ数値が出ることを想定していたが，今回の実験結果では，SIFT と SURF を組み合わせた場合と，それぞれを単一で使用した場合を比べると数値に差があり，SIFT と SURF を組み合わせた特徴量については提案方式において有効であると考えられる．

また実験 3 の結果により，全探索を行わずに組み合わせを提示する場合にでも，与えられた画像セットに対して評価関数が最大となるような画像特徴量を自動選択できていることがわかる．

今回の画像セットでは Purity が最大となる組み合わせが枝刈りの対象にならなかったが，画像セットによっては枝刈りの対象となることも考えられる．また，その場合には組み合わせる画像特徴量の最大数 n やランキングの上位から取り出す画像特徴量の数 m を適切に設定する必要がある．

3.6 おわりに

本稿では，Bag-of-Keypoints を用いて，画像を入力として画像特徴量の組み合わせを提示する方式を提案した．また，自動選択する画像特徴量としてキーポイント検出を 4 種，特徴量記述を 2 種設定したシステムを構築した．また実験により，構築したシステムに複数の学習画像セットを入力し，分類に有効な画像特徴量の組み合わせが出力されることを示した．このことから，提案方式の有効性を示した．

今後の課題として，今回の実験で構築したシステムの実験条件が出力に与える影響を調査することが挙げられる．提案したシステムは Visual-words 作成と画像特徴量の自動選択システムに分割でき，どちらにも実験条件が設定できる．Visual-words の作成においては画像の枚数やその画像の内容，Grid-sampling の間隔，Visual-words の数などを任意に決められる．

提案方式の組み合わせる画像特徴量の最大数 n やランキングの上位から取り出す画像特徴量の数 m を適切に設定する手法を考案することが挙げられる．今回の実験で用いた数値は暫定であり，設定する数値によっては最適な組み合わせの画像特徴量をユーザに提示できない可能性がある．

また，今回提案した組み合わせ探索の枝刈り方式以外の方式を実装することが挙げられる．機械学習の分野の特徴選択で用いられている相互情報量や χ 二乗検定などを提案方式でも扱えるように拡張することで，探索数を更に削減できる可能性がある．

今回は画像分類について，画像特徴量の組み合わせが実現できることを示した．本研究では画像特徴量を使用する目的として画像分類に着目したが，画像分類の他にも物体認識，類似画像検索などの目的が考えられる．今後の展望として，画像分類以外の目的について画像特徴量の組み合わせが可能な手法を考案し，画像特徴量を用いる様々な目的に対応できる方式を提案することが考えられる．

また，CNN を用いた画像分類との比較も検討する必要がある．

第4章　授業内発想支援システム: AI-Josyu

4.1　研究の背景

現代社会には実世界の人間の行動を感知するセンサーが数多く設置されており，多種多様で巨大なデータを簡単に取得することが可能となった．これらのセンサーの中には現実に起こる事象を画像，映像といったメディアデータとして感知し，リアルタイムにデータを生成し続けているものがある．一方で，これまでインターネットの文化が発展していく途中で，前述のようなセンサーによる自動感知ではなく，インターネットの歴史の中で自然と構築されてきた巨大なメディアコンテンツが存在する．現実の事象から自動生成されたメディアデータは人間側の無意識のうちに生成されているが，この自動生成されたメディアデータを対象としたデータ分析を行うために，レガシーシステムに散在するメディアコンテンツとの関係を自動で発見し，人間に提示する枠組みを整備する必要がある．

メディアデータとメディアコンテンツの相互接続には，どちらにも横断的に適用可能な，意味的メタデータを抽出する機構が必要不可欠である．この意味的メタデータを付与する機構は，音声認識や画像認識といった手法で実現することができ，近年では機械学習を用いた手法による発展がめざましい [25]．我々はこの意味的メタデータを付与する機構に適用可能な機構である Media-lexicon transformation operator($\mathcal{ML}$) を提案しており，画像や楽曲データに対して，それらメディアコンテンツの印象を表す語群を付与することが可能である [7, 26, 27]．これらの機構と実世界のセンサーを組み合わせることにより，自動的に実世界の事象を表すメタデータを収集することが可能となる．

本章では，前述の枠組みを備えたフレームワークであるメディアドリブンリアルタイムコンテンツマネージメントフレームワークを提案する．このフレームワークはメディアデータとメディアコンテンツの相互接続を実現するものである．このフレームワークは相互接続を実現するための 4 つのモジュールを備えている．“acquisition” モジュールは実世界からセンサーを通してメディアデータを獲得する．“extraction” モジュールはメディアデータを解析し，言葉として表された意味的メタデータを抽出する．“selection” モジュールは抽出されたメタデータの集合から，それらメタデータの背後に存在するコンテキストを選択する．このコンテキストは，レガシーシステムにアクセスするためのクエリーとなる．“retrieval” モジュールは，コンテキストに対応したメディアコンテンツをレガシーシステムから収集する．これら 4 つのモジュールにより，意味的メタデータを介した実世界のメディアデータとレガシーシステムに散在するなメディアコンテンツの相互接続が可能となる．このフレームワークによって構築されたシステムは自動的にメディアコンテンツを収集することが可能となる．

また，このフレームワークをベースとし，授業で発生するメディアデータを活用した授業

内発想支援システム *AI-Josyu* を提案する．さらにこの AI-Josyu に授業支援システムとしての機能を実現し，小中学校の先生方を対象に，授業の準備の負担を減らすためのシステムを構築する．このシステムは授業内で先生が発した言葉をリアルタイムに収集し，授業内容に関連したメディアコンテンツを即座に生徒に示すことができる．このシステムはインターネットブラウザ上で動作し，黒板に投影して使用する．画面を投影することにより，メディアコンテンツを事前に準備してハンドアウトとして配布する必要がなくなる．また，先生の発言内容に関連した単語を自動的に探索して提示し，生徒に発想支援を行うことができる．上記についての授業内発想支援システム AI-Josyu の具体的実装例と小学校における使用例を示す．

以下，2 節にて本章に関する関連研究，3 節にてメディアドリブンリアルタイムコンテンツマネージメントフレームワークの概要，4 節にて AI-Josyu の実装，5 節にて実使用を模した使用例，6 節にて小学校における実証実験，7 章にて本章の総括を述べる．

4.2 授業内発想支援システムの関連研究

音声波形を入力としてテキストを自動で書き起こす技術である音声認識は，近年機械学習の分野で大きく発展している [25]．機械学習による音声認識は，特定言語に特化したモデルが存在し，例えばインドネシア語に特化したモデルがある [28]．音声認識による授業の自動書き起こしが行われている例がある [29]．授業の自動書き起こしは教える側・教わる側双方に利点がある [30]．ビデオによる授業を配信しているようなケースでは，音声認識を使用してビデオにテキストデータを付与することによって，授業内容を検索できるようにしているものがある [30, 31, 32]．また，聴覚障害者の代わりにノートを取るノートテイクの代わりを自動書き起こしシステムが担う方法が提案・実装されている [33, 34, 35]．

Clicker は授業支援システムのひとつであり，授業内での学生の反応をリアルタイムに収集できるシステムである [36, 37]．学生の手元に操作可能なリモコンや，*Clicker* に対応しているスマートフォンアプリを用意し，学生側から授業に対して感じたことを情報として送るシステムになっている．

実世界と仮想世界間におけるセマンティックコンピューティングの実現は，これらの世界の相互接続のために重要なことである．駅での情報表示サービスのためのセマンティックコンピューティングによるランキング手法が提案されている [38]．

提案システムは，単に自動書き起こしを実現するのではなく，学生の発想を広げることに着目したシステムであるという点で，上記の自動書き起こしシステムと異なる．また，レガシーシステムに接続してメディアコンテンツを活用した発想支援システムである点が異なる．提案システムの大きな狙いは授業内容を保存し，同じような内容の授業を行う先生間で授業資料を共有するところにある．

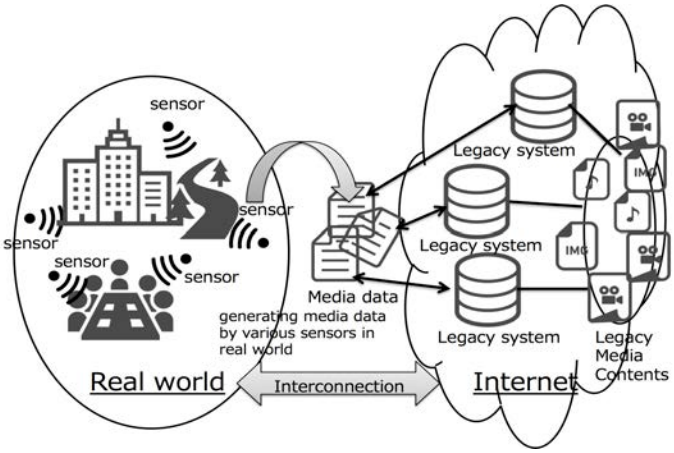


図 4.1: 実世界とインターネットのメディアコンテンツを介した相互接続

4.3 メディアドリブンリアルタイムコンテンツマネージメントフレームワークの概要

この節では，メディアドリブンリアルタイムコンテンツマネージメントフレームワークの概要について述べる．このフレームワークは実世界の事象から生成されたメディアデータとインターネット上に散在するメディアコンテンツを相互接続する機能を提供する．以下にフレームワークの概要と，フレームワークを構成する 4 つのモジュールの機能について述べる．

4.3.1 実世界とインターネットの相互接続

図 4.1 に実世界に設置されたセンサーとメディアコンテンツの相互接続の概要を示す．実世界には，様々な種類のセンサーが無数に設置されている．これらのセンサーは常時，メディアデータを生成し続けている．メディアデータは感知する事象に応じて画像，映像，音声などの様々な形式で保存される．これら生成されたメディアデータは，現実の事象を表現するデータである．

一方で，インターネット上のレガシーシステム上にはメディアコンテンツが散在している．レガシーシステムはそのようなメディアコンテンツを共有したり，検索したり，統合したりする機能を備えている．

このようなメディアコンテンツを利活用するためには，生成されたメディアデータとメディアコンテンツをシームレスに相互接続する必要がある．そのような相互接続を可能とするのが，以下に述べる 4 つのモジュールである．

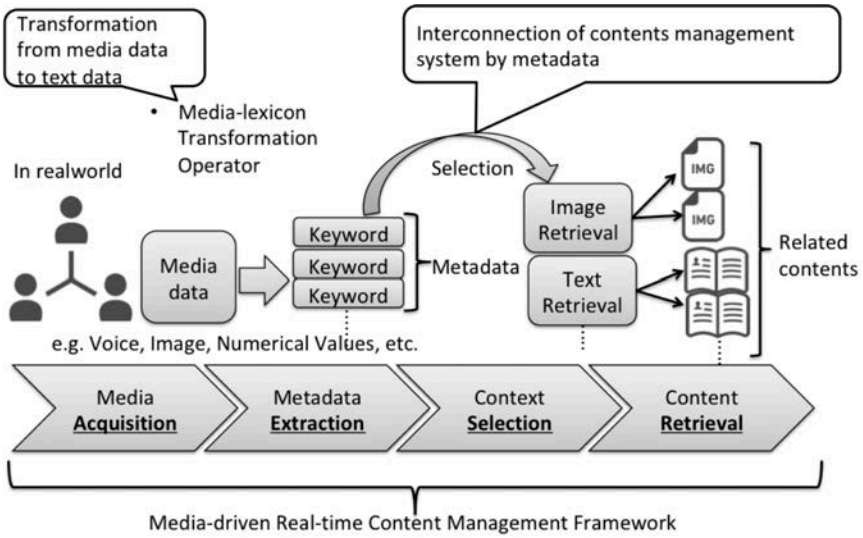


図 4.2: メディアドリブンリアルタイムコンテンツマネジメントフレームワークの概要図

4.3.2 メディアドリブンリアルタイムコンテンツマネジメントフレームワークを構成する 4 つのモジュール

図 4.2 はメディアドリブンリアルタイムコンテンツマネジメントフレームワークを構成する 4 つのモジュールの概要図である．4 つのモジュールの実現により，メディアデータとメディアコンテンツを相互に接続することが可能となる．

このメディアドリブンリアルタイムコンテンツマネジメントフレームワークは “acquisition” モジュール，“extraction” モジュール，“selection” モジュール，“retrieval” モジュールで構成される．以下に詳細を示す．

(1) Media Acquisition
acquisition モジュールは，実世界の事象をセンサーを通してメディアデータとして獲得するモジュールである．メディアデータは実世界に配置された多数のセンサーがリアルタイムに生成する．それぞれのメディアデータは，画像，音声，映像など，様々な形式を取る．

例としては，マイクによる音声の取得や，カメラによる画像・映像の撮影が当たる．このモジュールにより，実世界の事象をデジタル化する．

(2) Metadata Extraction
extraction モジュールは，メディアデータから意味的メタデータを抽出するモジュールで

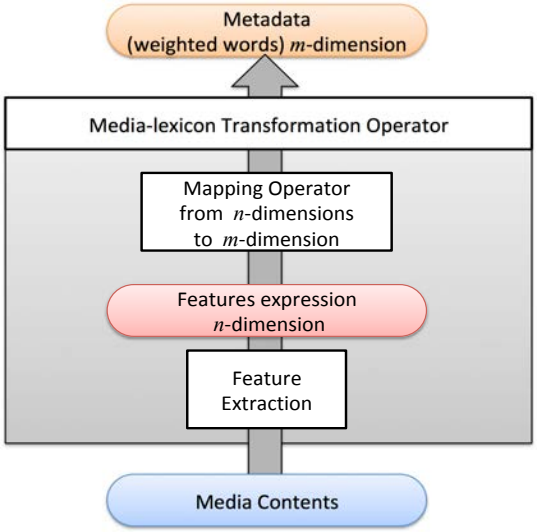


図 4.3: Media-Lexicon transformation operator

ある．メタデータは言葉として記述され，メディアデータに実世界の事象の説明を付与する．言葉として記述することにより，レガシーシステムへのアクセスが可能となる．

このモジュールの例としては，機械学習で発展した画像認識・音声認識などといった手法が挙げられる [25]．これらはメディアデータに意味的メタデータを付与することが可能である．例えば音声認識であれば音声データからテキストの書き起こしを自動的に抽出可能である．また，画像認識を用いれば，画像に存在している物体の名前を付与したり，キャプショニングによって画像そのものを説明する文章を生成したりすることが可能である．このモジュールには北川らが提案している $\mathcal{ML}$ を適用することも可能である [7, 26, 27]．$\mathcal{ML}$ はメディアコンテンツに対し，そのメディアコンテンツから受ける印象を重み付き単語群として抽出する作用素であるが，メディアデータにも同様に適用可能である．$\mathcal{ML}$ は以下のように定義される．

$$
\begin{aligned}
\mathcal{ML}(Md) &: Md \mapsto Ws, \\
\text{ただし，} & \\
\mathcal{ML} &: \text{Media-Lexicon transformation operator}, \\
Md &: \text{メディアコンテンツ}, \\
Ws &: \text{重み付き単語群}.
\end{aligned}
$$

図 4.3 は $\mathcal{ML}$ の概要図である．$\mathcal{ML}$ は，メディアコンテンツから，言葉のメタデータ

空間を構成する重み付き印象語群によって特徴づけられた m 次元ベクトルを抽出する作用素である．$\mathcal{ML}$ による言葉のメタデータを抽出するステップは 2 つから成り，1 つ目が特徴抽出，2 つ目は言葉のメタデータ空間へのマッピングである．特徴抽出では，メディアコンテンツからそのメディアコンテンツをよく表現する特徴を抽出する．この特徴には，それぞれのメディアコンテンツの専門家が定義したものを使用する．この特徴抽出により，n 次元のベクトルを抽出する．言葉のメタデータ空間へのマッピングは，メディアコンテンツの専門家によって明らかにされたメディアの特徴と言葉の関係によってマッピングのための作用素を使用する．これにより，n 次元のベクトルを m 次元の重み付き印象語群ベクトルにマッピングする．

(3) Context Selection
selection モジュールは，抽出された言葉のメタデータからコンテキストとなる言葉を選択するモジュールである．コンテキストとなる言葉は，言葉のメタデータ空間上に存在する単語から選択される．このモジュールで選択されたコンテキストは，レガシーシステムにアクセスするためのクエリーとして用いる．

このモジュールは，単語に関するセマンティックコンピューティングを行う部分を含む．言葉に関する意味的な類似度や関係，距離を計量できるモデルを使用してコンテキストを選択する．

(4) Content Retrieval
retrieval モジュールは，コンテキストに対応したメディアコンテンツを収集するモジュールである．メディアコンテンツはインターネット上に散在しており，レガシーシステムを通して収集される．

これらのモジュールによって構成されたフレームワークの実現により，実世界のメディアデータとレガシーシステム上のメディアコンテンツを意味的かつ相互に接続することが可能となる．意味的なメタデータとして表現された実世界の事象に，レガシーシステムから意味的かつ自動的に収集されたレガシーなメディアコンテンツを対応付けることが可能となる．

4.4 メディアドリブンリアルタイムコンテンツマネージメントフレームワークによる授業内発想支援システムの実装

この節では，授業内発想支援システム *AI-Josyu* の実装について述べる．このシステムはメディアドリブンリアルタイムコンテンツマネージメントフレームワークをベースに実装されている．システムは授業中の先生の発言をリアルタイムにマイクで取得し，その発言に関連しているメディアコンテンツをインターネットから自動で収集する．システムは Web ブラウザ上で動作し，Web ブラウザで表示している内容をプロジェクターによって黒板上に投影して使用する．これにより，メディアコンテンツをあらかじめ収集する必要がなくなる．また，メディアコンテンツを表示することによって，生徒への発想支援を行うことが可能である．

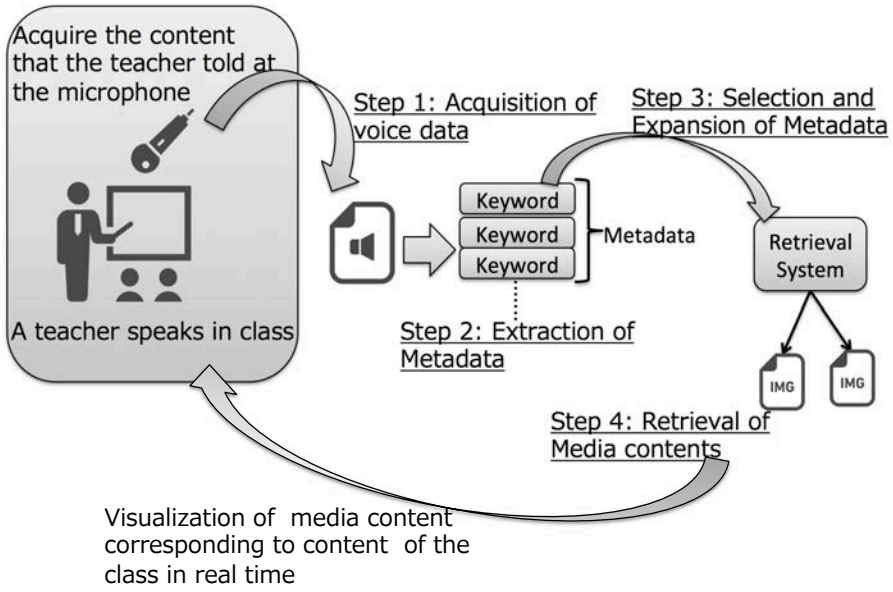


図 4.4: AI-Josyu の動作

図 4.4 はメディアドリブンリアルタイムコンテンツマネージメントフレームワークと対応付けて AI-Josyu の動作を説明している図である．各モジュールの具体的な実装を以下に述べる．

Step 1: 先生の発言の獲得 (Acquisition module)
先生の発言を録音できるように，先生の襟元にマイクを取り付ける．録音された発言データをシステムに格納する．

Step 2: キーワードの抽出 (Extraction module)
録音された発現データをテキストデータとして書き起こし，そのテキストデータからキーワードを自動的に抽出する．このモジュールは，$\mathcal{ML}$ と音声認識によって実装されている．AI-Josyu においては，Google Chrome に実装されている Web Speech API を使用してテキストを自動で書き起こす．

Step 3: 授業内重要単語度の計算と関連語の表示 (Selection module)
AI-Josyu におけるコンテキストとは，授業内重要単語である．先生は extraction モジュールで抽出されたキーワードから自分で選択することも可能であるが，システムがキーワードの重要度を自動的に算出しており，それを参考に選択することができる．この授業内単語重要度の算出方法については後述する．また，システムは先生が選択したコンテキストに関連している言葉を自動的に抽出し，表示する．

この授業内単語重要度の計算と関連語の表示には，単語の分散表現によって言葉の類似

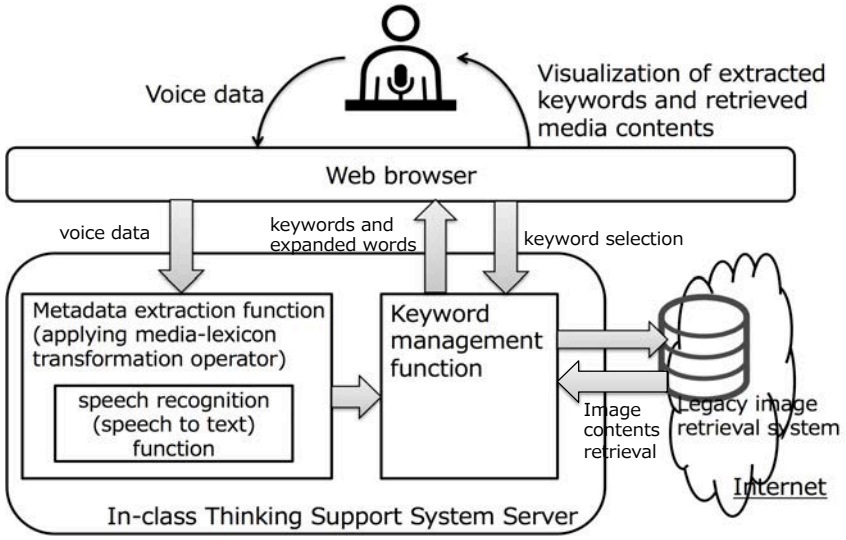


図 4.5: AI-Josyu の構成

度を計量できる Word2Vec モデル [39, 40, 41] を使用する．

Step 4: 画像の収集 (Retrieval module)
AI-Josyu は，先生が選択したコンテキストに関連した画像をレガシーシステムを経由して収集する．AI-Josyu においてはレガシーなシステムとして Google の画像検索を使用する．

これらのモジュールによって収集された画像は，黒板に投影された画面上に表示される．これにより，メディアコンテンツを含む資料を事前に作成する必要がなくなる．また，授業に関連したメディアコンテンツをリアルタイムに学生に示すことによって，学生の発想支援を促す．

図 4.5 は AI-Josyu の構成である．AI-Josyu は Web ブラウザ上で動作し，その画面を黒板に投影して使用する．先生は AI-Josyu が動作しているコンピュータに接続されたマイクを使用して，授業内の発言をリアルタイムにシステムに入力する．メタデータ抽出機能は音声認識を備えており，先生の発言をテキストに自動で書き起こす．書き起こされたテキストは，キーワードマネージメント機能に入力され，テキストの解析によって，授業内コンテキストを設定したり，キーワードに類似した単語を選択する．授業内コンテキストを使用してレガシーシステムにアクセスすることにより，授業に関連した画像データを収集する．これらは AI-Josyu が自動で行うため，先生は発言した内容に関連したメディアコンテンツを即座に表示することができる．

授業内で発言されたキーワードに類似している単語を選択するために，Word2Vec モデルを用いる．Word2Vec モデルは単語を分散表現のベクトルに変換することができ，ベクトルの類似度を計量することによって単語の類似度を計量できるモデルである．今回使用した Word2Vec モデルは，Wikipedia の日本語記事を学習したプリトレインモデルである．

4.5 授業内単語重要度の算出方法

本節では，キーワードマネージメント機能が行う授業内単語重要度の算出方法について述べる．提案システムにおいて “aqcuisition” モジュールはリアルタイムで先生の発言を獲得し続けるが，ある程度の長さの音声データを獲得した段階でそのデータを “extraction” モジュールに送る．このデータを送るタイミングについては使用する音声認識 API やその実装方法に依存する．今回構築したシステムでは音声認識 API の仕様に則り，発言に 2〜3 秒の沈黙があったときに “extraction” モジュールにデータを送るようになっている．

授業内単語重要度の算出にあたり，あらかじめ設定しておく必要のあるパラメータがいくつか存在するため，以下に列挙しておく．

単語間の関連の有無を決定するための閾値 ϵ

今回の授業内単語重要度においては，ある単語に着目したとき，その周辺に存在する単語と関連があるかどうかを判断する必要がある．そのため，単語間関連度を計量できるモデルにおいて，ある単語とある単語について類似度または単語間距離が定義されているとき，本システム内において関連がある単語であると判定する閾値 ϵ を定める．

Word2Vec モデルにおいては，ある単語を表すベクトル同士の類似度はコサイン類似度で定義されているため，単語間の関連の度合いは 0 から 1 の値を取り，1 に近づくほど類似していることになる．提案システムにおいては $\epsilon = 0.3$ と定め，0.3 よりも値が大きい 2 つの単語は関連があるとみなすことにする．

授業内コンテキストを抽出する時間 t_{int}

授業内単語重要度は，授業内コンテキストと発言の関連度を使用して算出される．このとき，授業内コンテキストは基本的に，その授業開始時点から授業内単語重要度を算出する発言が行われるまでのすべての発言を参照して決定する．しかし，実使用の場面においては，同じ授業時間内で複数のテーマが展開されるなど，授業内の発言すべてを参照すべきでない場合も存在する．そのため，先生の発言が行われたとき，どこまで遡って授業内コンテキストを参照するかを決定するための t_{int} を定める．

提案システムにおいては，協力企業と調整を行い，$t_{\mathsf{int}} = 300\mathsf{s}$ と定め，300 秒まで遡って授業内コンテキストを設定することとした．

出現頻度スコアに対する係数 α と関連度スコアに対する係数 β

最終的な授業内単語重要度は，出現頻度に応じて決まるスコアと，関連度に応じて決まるスコアの重み付きの足し合わせで求める．

提案システムにおいては，協力企業と調整を行い，$\alpha = 0.4, \beta = 0.6$ とした．関連度スコアのほうが比率が大きいため，この場合は直近の発言よりも過去の発言を重要視するモデルとなる．

提案システムにおける授業内単語重要度は以下の手順で算出される．以下の手順は，“extraction” モジュールへのデータ送信が 1 回行われるごとに 1 回実行される．

Step 1: 授業内コンテキストの決定
先生の発言が行われた時点から t_{int} まで遡った範囲の発言を参照し，その範囲に存在する単語を形態素解析によって抽出する．これによって得られた単語群を授業内コンテキストとする．

Step 2: 授業内コンテキストにおける出現頻度スコアの計算
授業内コンテキストに関して発言単語をカウントし，各単語の出現頻度スコアとする．ただし，授業内コンテキストで一番発言数の多い単語の発言数で正規化し，0 から 1 の範囲の値を取るようにする．

Step 3: 授業内コンテキストの単語間関連度における関連度スコアの計算
授業内コンテキストから任意の 1 つの単語を取り出す．この取り出した単語以外のすべての授業内コンテキスト内単語とペアを作り，Word2Vec にて単語間類似度を計量する．このとき，閾値 ϵ 未満であるとき，および同じ単語であるときは計量を行わずスキップする．ϵ 以上の単語のペアについて関連度の平均を取り，取り出した単語の関連度スコアとする．

Step 4: 授業内単語重要度の計算
Step 2 の出現頻度スコアと Step 3 の関連度スコアについて，α と β の重み付き足し合わせによって最終的な授業内単語重要度とする．

4.6 AI-Josyu の使用例

本節では，実使用を模した AI-Josyu の使用例を，実際の AI-Josyu の画面表示とともに述べる．

ここでは歴史の授業で豊臣秀吉について教えるという簡単なケースで説明を行う．ここに載せている画像は Web ブラウザのキャプチャであるが，実際には黒板上に投影されており，先生はチョークで説明を書き足すことができる．

図 4.6 に acquisition モジュールにおける音声の獲得の様子を示す．Web ブラウザ上で AI-Josyu を起動すると，黒い画面といくつかの機能を持ったボタンが表示される．マイクおよびブラウザの設定を音声を取得できる状態にすると，システムは音声の取得と認識を自動で開始する．

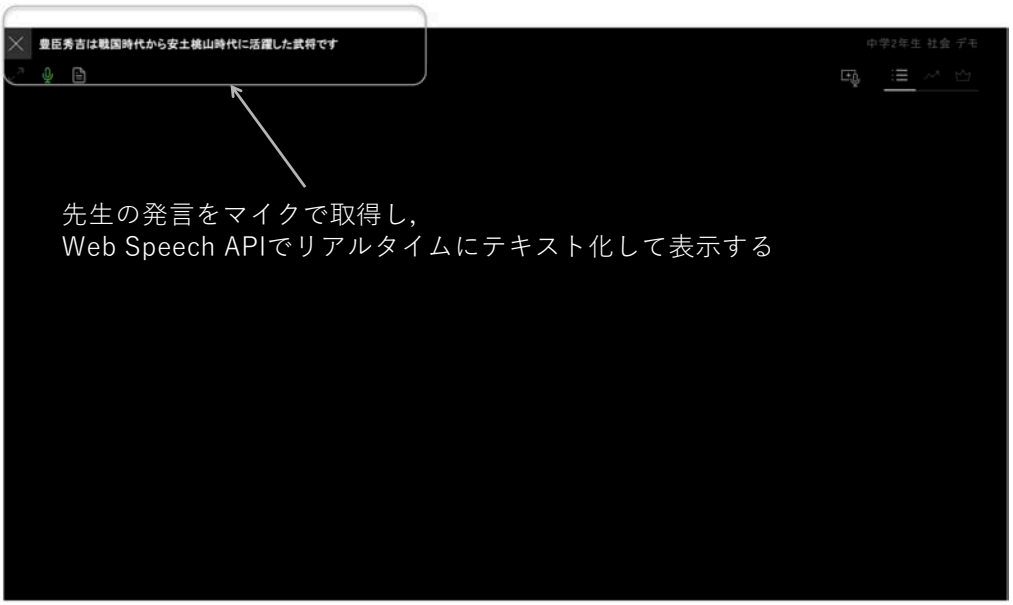


図 4.6: AI-Josyu の acquisition モジュールによる音声の獲得

ここでは，先生が「豊臣秀吉は戦国時代から安土桃山時代に活躍した武将です」と発言した．システムはリアルタイムに音声を取得し，自動でテキストを書き起こし，左上にリアルタイムで表示する．

図 4.7 は extraction モジュールの様子である．システムは書き起こされたテキストから意味的メタデータとなる単語を抽出し，右側に表示する．このとき，キーワードマネージメント機能は，その授業内での過去の発言を参照しながら，抽出された単語に重要度を与える．その重要度を参照し，重要度の高いものほど上になるように表示する．

この例では，先程の先生の発言から「武将」「活躍」「安土桃山時代」「戦国時代」「豊臣秀吉」の 5 つの単語が抽出された．この順番は，キーワードマネージメント機能が定めた重要度に基づくものであり，この発言においては「武将」が最も重要な単語と判断されている．先生は，この単語リストに表示された単語をドラッグアンドドロップすることによって，中央の空きスペースに単語を配置することができる．

図 4.8 は選択した単語の関連語ネットワークを表示する部分である．ドラッグアンドドロップで配置された単語をクリックすると，配置された単語に関連している単語をネットワークの形で表示する．この例では，選択した「豊臣秀吉」と関連度の高い単語として「九州征伐」「天正」「豊臣秀次」「小田原征伐」「秀吉」が表示されている．

図 4.9 は retrieval モジュールによるメディアコンテンツの収集の様子である．描画されたネットワークには収集したメディアコンテンツを表示するボタンがあり，そのボタンをクリックすることによって，単語との関連度の高いメディアコンテンツを表示する．現在の AI-Josyu

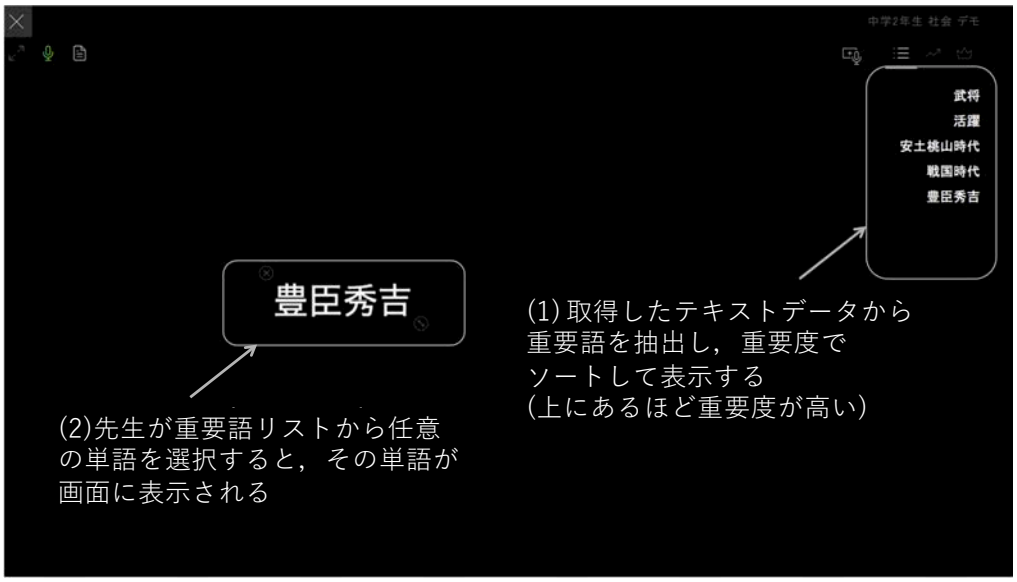


図 4.7: AI-Josyu の extraction モジュールによる自動書き起こしとキーワード抽出と，selection モジュールによるコンテキスト選択

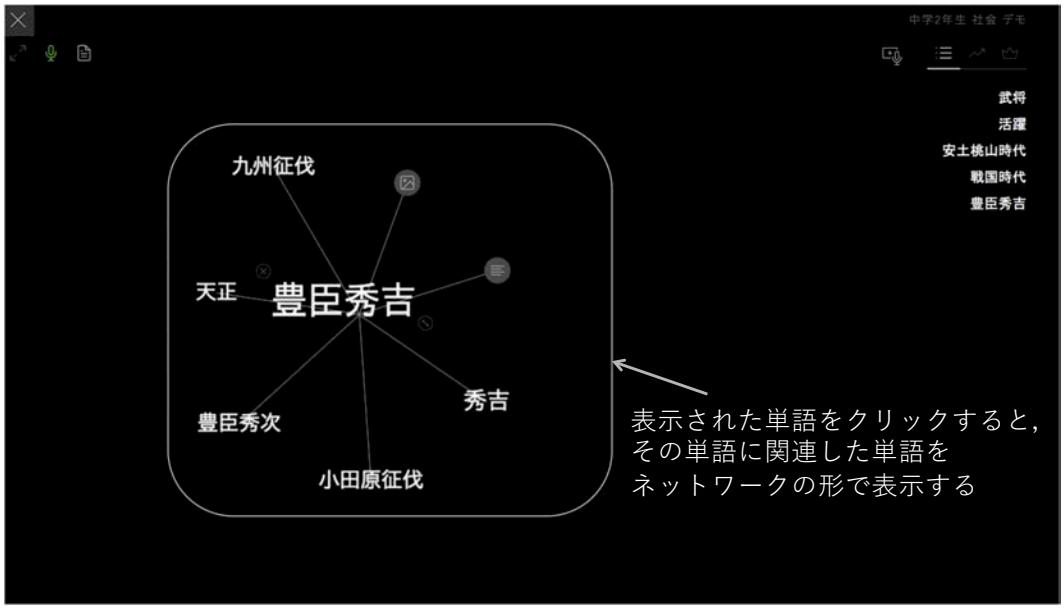


図 4.8: 関連語ネットワークの表示

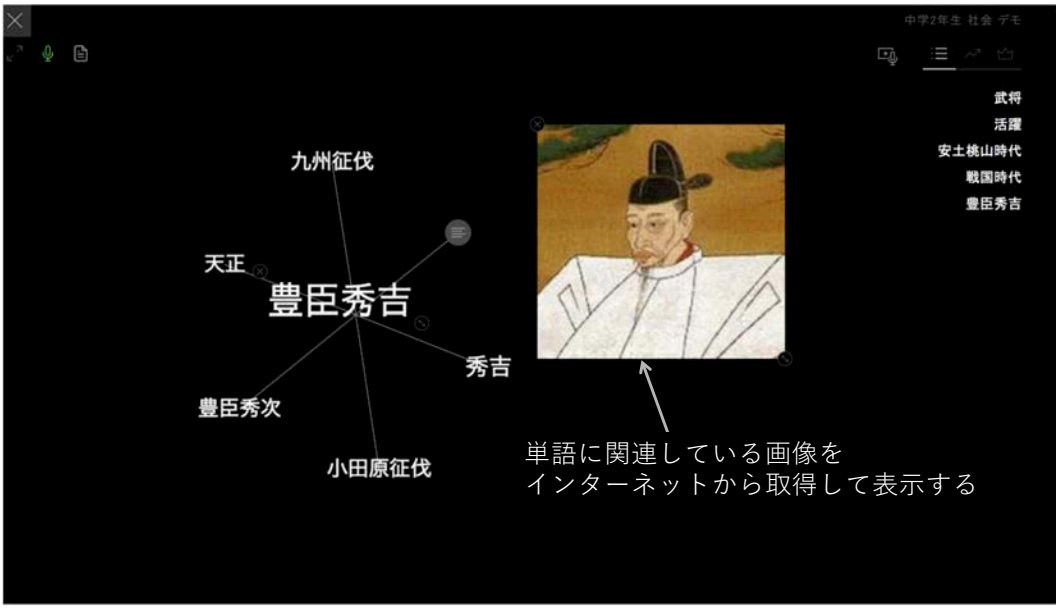


図 4.9: AI-Josyu の retrieval モジュールによるメディアコンテンツの収集と表示

では，画像を表示することが可能である．

この例では「豊臣秀吉」と関連度の高い画像をシステムが自動で収集しているため，画像表示ボタンを押すことによって豊臣秀吉の肖像画を即座に表示することができる．また，必要に応じて，最も関連度の高い画像だけではなく，複数の候補から選択して表示することも可能である．

4.7 小学校における実証実験

本節では授業内発想支援システムである AI-Josyu を実際に授業で使用していただいた様子について述べる．東京都日野市にある平山小学校に協力していただき，理科 (地層)，歴史 (飛鳥時代)，理科 (人間のからだ) の 3 つの授業で使用していただいた．また，使用してみての感想を伺った．

4.7.1 実験環境

図 4.10 は，AI-Josyu を使用しているときの様子の写真である．黒板の前に AI-Josyu をブラウザで開いた状態のノートパソコンと，それを投影するためのプロジェクターを用意する．先生の襟元にはワイヤレスマイクが装着され，発言した内容がノートパソコンに送られる．

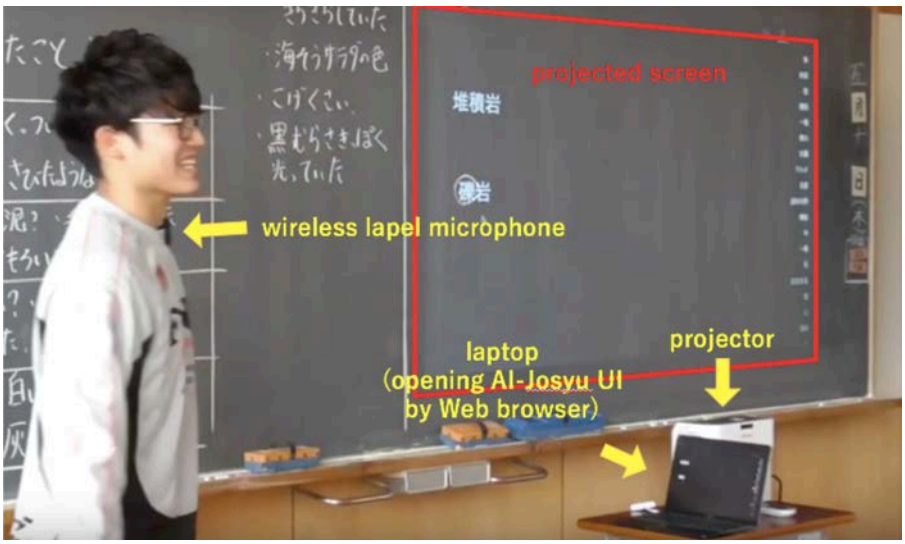


図 4.10: AI-Josyu を使用するための環境

それぞれの授業を 30 人ほどの生徒が受講している．生徒側にマイクは取り付けられていないため，先生の音声のみを取得するようになっている．

4.7.2 理科 (地層) における活用

図 4.11 は理科 (地層) の授業において，レガシーな画像検索システムを活用しながら講義をしている様子である．この授業においては右側の単語リストを黒板に投影せずに使用しているが，先生が発言した内容は音声認識を通して自動でテキスト化され，キーワードマネージメント機能によって重要であると思われる単語がリスト化されている．この先生はそのキーワードリストから堆積岩，礫岩，砂岩の 3 つを選択し，システムが収集した礫岩と砂岩の画像を表示して使用している．先生はこの画像について事前に準備することなく，発言した内容に応じて自動的にシステムが収集した画像を選択するのみでメディアコンテンツを使用した講義を行うことができた．岩石の画像を見せることにより，さらに生徒の理解が深まったと考えられる．

この画像収集はメディアドリブンリアルタイムコンテンツマネージメントフレームワークによってリアルタイムかつ自動的に行われている．このような仕組みにより，レガシーシステムを通して，巨大なメディアコンテンツを簡単に活用できるようになる．

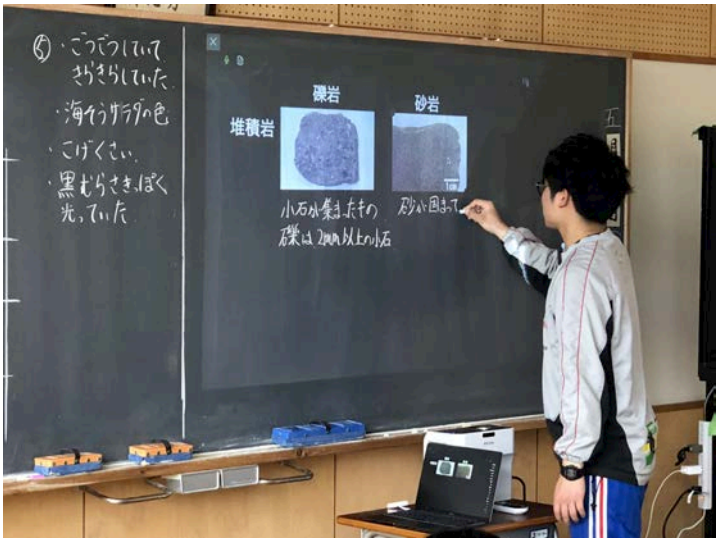


図 4.11: 理科 (地層) で AI-Josyu を活用している様子

4.7.3 歴史 (飛鳥時代) における活用

図 4.12 は歴史 (飛鳥時代) の授業において，関連語ネットワークを使用した「聖徳太子」の理解支援を行っている様子である．AI-Josyu が自動で書き起こし，抽出された重要語が黒板の右側にリストアップされている．これらのキーワードは授業全体を通したコンテキストに応じて重要度が決定されている．ここでは，先生は「聖徳太子」という単語をメインのキーワードとして選択し，中央に配置した．キーワードマネージメント機能が提供する関連語ネットワーク機能を使用して，聖徳太子に関連する単語を表示し，授業に活用している．この関連語ネットワークには，聖徳太子が行ったことや聖徳太子に関係している人物が関連語として表示されている．

さらに，先生はその関連語を深掘りして「十七条憲法」を再度選択し，関連語ネットワークを表示させた．これらは生徒に対して，キーワードに対する理解を深めるとともに，興味を広げさせる効果があると考えられる．

図 4.13 は，投影されたネットワークに先生がチョークで情報を書き加えて活用している様子である．画面を黒くしているため，チョークによって上書きして使用することが可能である．

4.7.4 理科 (人間のからだ) における活用

図 4.14 は理科 (人間のからだ) の授業において，人間の臓器の関係を説明するために AI-Josyu を活用している様子である．レガシーなシステムから収集された人間の体の画像を表示し，理

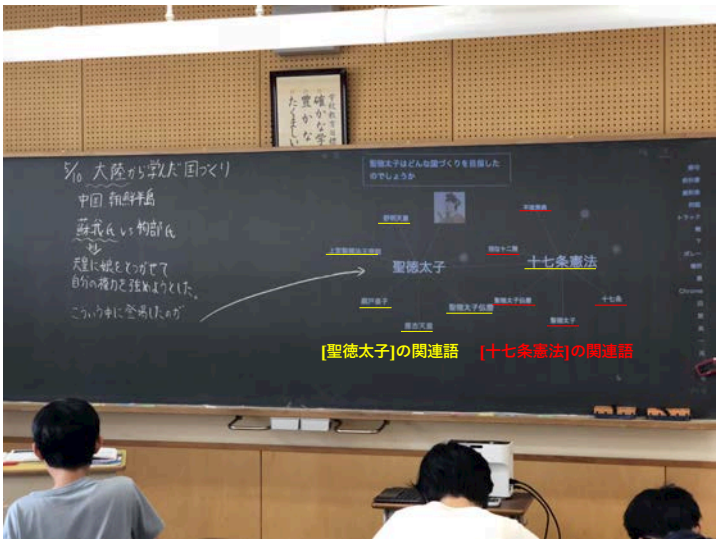


図 4.12: 歴史 (飛鳥時代) で AI-Josyu を活用している様子

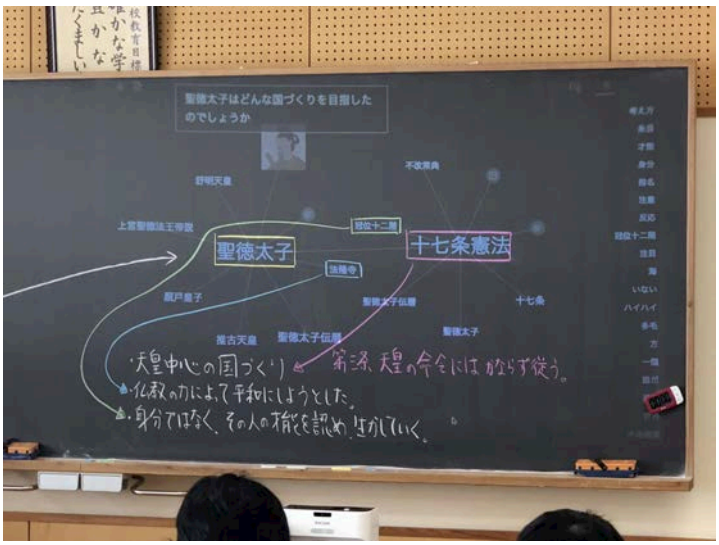


図 4.13: チョークによる上書きが可能である様子

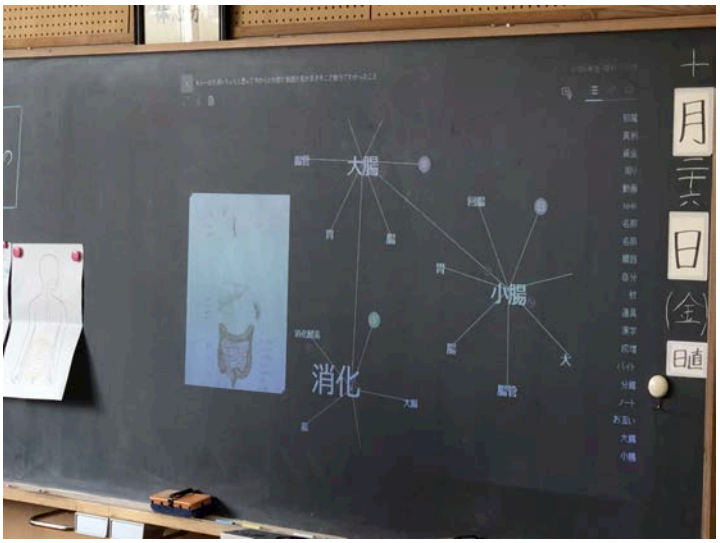


図 4.14: 理科 (人間のからだ) で AI-Josyu を活用している様子

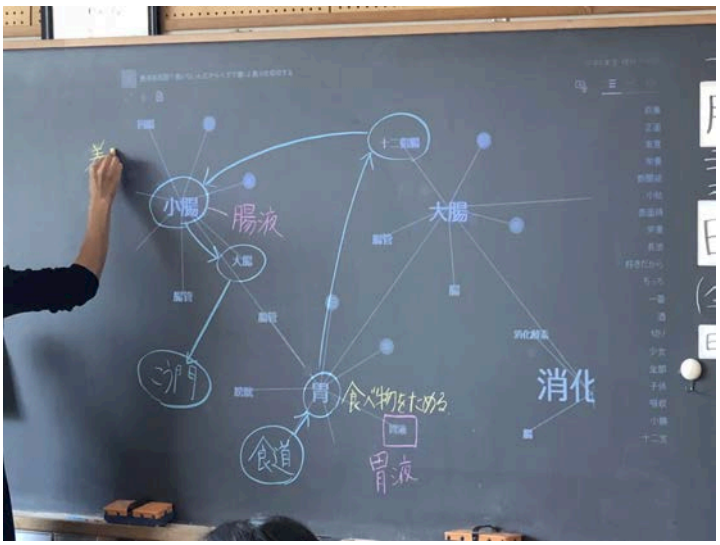


図 4.15: 関連語ネットワークと矢印の上書きによる関係性の図示

解支援を行っている．また，関連語ネットワークを活用することによって，人間の臓器の関係を図示している．図 4.15 では，さらに関連語ネットワークに表示された単語に対して矢印を使用したり，説明を書き加える関係性の図示を行うことで，より理解が深まると考えられる．

4.7.5 AI-Josyu に関するインタビューと考察

今回のシステムを使用した感想を先生方に伺った．主に，使いやすさと，期待しているような性能であるかを伺っている．

まず，使いやすさについて,「システムは簡単に使えるように作られていると思う．選択したキーワードがなぜ重要なのか，可視化してくれるのが良い．自動的なキーワード抽出や関連語ネットワークによって，生徒が自分で学習のゴールを設定し，理解しようという意欲が見られる．自分のアイディアと関連語を紐づけて考えることができている．特別な支援を必要とする生徒にとって，視覚的に表示される点が利点である.」と述べている．

性能について尋ねると,「(最初のクリックでは関連語ネットワークは 5 つまでしか単語が表示されないため，) 表示される 5 つの関連語は，期待している結果にならないこともあった．しかし，関連度が 6 番目以降の関連語を選択してネットワークに加えることができる機能があったので，その点については満足している．やはり，先生の発言だけではなく，生徒の発言も取得して可視化されると良い．それによって，より理解が深まると思う.」と述べている．

4.8 結論

この章ではメディアドリブンリアルタイムコンテンツマネージメントフレームワークの提案と，それをベースとする授業内発想支援システム *AI-Josyu* を提案した．このフレームワークは実世界の事象から生成されたメディアデータとレガシーシステムに散在するメディアコンテンツを相互に接続することができる．これにより，実世界の事象に対してシームレスに意味的メタデータを付与することが可能となった．

また，*AI-Josyu* のプロトタイプシステムを実装し，実際の授業で使用していただいた．このシステムは先生の発言をリアルタイムに録音し，授業の内容に関連したメディアコンテンツをインターネット上から収集して表示することができる．このシステムは黒板に投影して使うため，先生の授業準備の負担を軽減することができる．メディアコンテンツや関連語ネットワークの表示により，授業に対する生徒の理解支援に役立つシステムとなっている．

今後の改善点としては，現行のプロトタイプシステムに翻訳機能を搭載し，複数言語間でも相互に使用できるシステムとする点が挙げられる．また，現状ではレガシーシステムとして画像検索システムを使用しているが，画像以外の映像や音声といった形式を取り扱うレガシーなシステムとの接続を行えるシステムに改善することができる．また，acquisition モジュールに相当する部分についても，音声の獲得のみではなく，カメラを用いた映像の解析やスマートフォンアプリなどを用いた生徒の反応の獲得などの機能を追加し，より理解が深まる授業を行えるシステムにする点が挙げられる．

第5章　CM字コンテ好感度予測システム: CREATIVE BRAIN

5.1　研究の背景

今日，我々は日常的に数多くのテレビCMを目にしている．近年では，インターネットによる広告市場の規模が増加してきていると報告されているが，テレビCMはその規模や役割から，未だに広告における中心的役割を担っているといえる．2018年度の調査[1]では，関東地方だけでも2,463社の7,098種類の商品・サービスについて15,168作品ものテレビCMが放送されており，また総放送回数は1,532,499回に上るとの報告がある[42]．この数多くのテレビCMは，広告主に当たる企業から視聴者に対して，何かしらのメッセージや印象を伝える役割を持っている．そのため，テレビCM作品は，視聴者に覚えてもらえるようなインパクトがあるか，また視聴者に対して企業の意図が明確に伝わるかを重視する．そのようなテレビCM作品の製作は，CMクリエイターの感覚に頼る部分が大きく存在する．

前述のような背景から，放送されているテレビCMについて視聴者が受けている印象を把握するため，テレビCMの大規模なモニター調査を行っている企業が存在する．この調査結果は新しいテレビCMを製作する際に活用できるデータであると考えられる．しかしながらこれまでは，CMクリエイターが作成したテレビCMの放送後に，そのテレビCMの出来の良さを評価するためにしか利用されていなかった．

また，これまで，様々なメディアコンテンツを対象として，そのメディアコンテンツの持つ印象を言葉のメタデータとして自動的に抽出する仕組みであるMedia-lexicon Transformation Operator($\mathcal{ML}$)[43]について研究が行われている．これまでに提案されてきた$\mathcal{ML}$は，対象となるメディアコンテンツに対応する心理学などの研究成果を用いることにより実現されてきた．$\mathcal{ML}$によって抽出されるメディアコンテンツの持つ印象を“印象メタデータ”と呼ぶ．具体的事例として，我々はこれまでに，楽曲を対象とした$\mathcal{ML}$[5, 6, 26]によって楽曲の印象メタデータ抽出を実現したり，画像の色を対象とした$\mathcal{ML}$[7]によって色の印象メタデータ抽出を実現したり，“音相”と呼ばれる言葉の発音情報を対象とした$\mathcal{ML}$[44]による言葉の発音の印象メタデータの抽出を行ったりしてきた．また，これらの$\mathcal{ML}$によって，印象を表現する言葉のメタデータによって異種メディアコンテンツを連結し，例えば楽曲を入力として画像を検索するなど，異種メディアコンテンツ間での検索システムなどといったコンテンツマネジメントシステムを実現することが可能となる．

そこで本章では，テレビCMの印象を自動抽出する$\mathcal{ML}$の構成方式を示す．この$\mathcal{ML}$は，

[1]集計期間は2017年10月20日から2018年10月19日まで.

テレビ CM 作品のセリフやキャッチコピー，説明文に含まれる単語から，テレビ CM の印象メタデータを自動抽出する．この $\mathcal{ML}$ の実現によって，テキストデータとして表現された字コンテからそのテレビ CM が視聴者に与える印象を推定できるため，CM クリエイターは，実際の映像作品を製作する前の段階で，複数の字コンテを比較して CM に期待している印象に近いものを選ぶといったことや，期待する印象に近づくように字コンテを修正することが行えるようになる．

これまでの $\mathcal{ML}$ では，対応するメディアコンテンツの専門家が行った既存の研究成果が用いられてきたが，本章で述べる CM の $\mathcal{ML}$ は専門家の研究成果ではなく，過去に放送されたテレビ CM をテキストで表現した字コンテと，テレビ CM の印象に関する大規模モニター調査データを用いて構成することを特徴とする．$\mathcal{ML}$ をデータから構成する本方式は，これまで専門家の研究が十分でなかった分野のメディアコンテンツに対しても $\mathcal{ML}$ を構成することが可能となる方式である．そのため，これまでよりも広い分野のメディアコンテンツに対応したコンテンツマネジメントシステムに応用することができる．

本手法で用いる大規模モニター調査データは株式会社東京企画（CM 総合研究所）が行っている CM 好感度[2]調査アンケートを用いる．

本章の構成は以下の通りである．2 節では，関連研究を示す．3 節では，$\mathcal{ML}$ について詳細に述べる．4 節では，3 節で述べた $\mathcal{ML}$ の適用事例として，テレビ CM を対象とした $\mathcal{ML}$ の構成方式について述べる．5 節では，実験により，本方式の有効性について検証する．6 節では，本稿のまとめを述べる．

5.2 関連研究

本章の提案手法および提案システムについて，その前段となる好感要因の推定手法と可視化手法の概要を述べたものが報告されている [45]．本章は，好感要因の推定手法の具体的な構成・計算方法を提案し，その性能評価について検証するものである．

テレビ CM に関する研究の一つとして，テレビ CM の実態分析 [46] が挙げられる．この分析では，1 週間に放送された CM のうち，食品・飲料品に関する CM に関して，その放送回数などの情報量，栄養や品質などの機能的ベネフィット，強調表示や景品の有無などの心理的ベネフィットを取り上げて行われている．また，テレビ CM の印象に関する研究としては，テレビ CM に出演しているキャラクターやタレントが認知や評価に及ぼす影響度を調査したもの [47] や，特定のテレビ CM について，自由記述形式の感想文をテキストマイニングにより調査したもの [48] が報告されている．

上記の研究は，放送されたテレビ CM に関する調査結果やその方法を提案していたが，本章で我々が提案する手法は，実際に放映される前のテレビ CM について，その好感度を推定できるという点で異なる．

また，本章では，テレビ CM を対象とした Media-lexicon Transformation Operator($\mathcal{ML}$) [5, 6, 26] の構成方式であり，テレビ CM の好感度の定量的な評価方法である．また，テレビ CM

[2]「CM 好感度」は株式会社東京企画（CM 総合研究所）の登録商標である．

を表すデータとして，出演者のセリフ等を書き起こしたテキストデータを用いるため，文章に対する印象を抽出する手法としても位置付けることができる．

文章に対する印象抽出の研究の 1 つとして，感情極性抽出 [49] が挙げられる．感情極性とは，それぞれの単語が持つ，“人間に良い印象を与える”（陽性）か，“悪い印象を与える”（陰性）かを表す二値変数，あるいはそれら二値に，“印象を与えない” を加えた三値変数であり，例えば “美しい” はポジティブな印象を与え，“汚い” はネガティブな印象を与えるといったものである．この感情極性の抽出には，インターネット上の語彙から既知の感情極性を持つ単語との共起性を計算する手法 [50] がある．感情極性の抽出には，各単語ごとの感情極性が示されたものである極性辞書が必要となる．この極性辞書の作成には人手で感情極性を付与したもの [51]，国語辞典の語釈文からブートストラップ的に獲得したもの [52]，WordNet[53] から語彙ネットワークを構築し，ネットワーク上での最短距離を用いて感情極性を決定するもの [54] が提案されている．

本稿で提案する我々の手法は，上記の感情極性抽出のように，“良い印象” または “悪い印象” という 1 種類の変数で表現される印象ではなく，大規模モニター調査データで得られる複数種類の印象について自動抽出する仕組みという点で，上記手法とは異なる．また，感情極性抽出では離散値での抽出であったが，本手法では連続値による抽出である点でも異なっている．

5.3 メディアコンテンツを対象とした $\mathcal{ML}$ の構成

本節では，メディアコンテンツから人間が受ける印象を言葉として自動的に抽出する方式である Media-lexicon Transformation Operator($\mathcal{ML}$)[43] を示し，また，その構成方式について述べる．Media-lexicon Transformation Operator($\mathcal{ML}$) は，対象とするメディアコンテンツの専門家による研究や評論，統計などを用いることにより，人間がそのメディアコンテンツから受ける印象を表す言葉のメタデータの抽出を実現する機構である．

$\mathcal{ML}$ は言葉同士の相関を計量する機構とセットで考案されたものである．様々な種類のメディアコンテンツの印象を言葉という統一的なメタデータで表現し，さらにそれら印象メタデータ同士の関係を計量する機構と組み合わせて，メディアコンテンツの分野を跨いだ統一的な操作を行うことを目指している．

$\mathcal{ML}$ は一般に次のように表される．

$$\mathcal{ML}(Md) : Md \rightarrow Ws.$$

$$(Md : \text{メディアコンテンツ}, Ws : \text{(重み付き) 印象語群})$$

$\mathcal{ML}$ による印象メタデータ抽出は，メディアコンテンツから特徴量への変換を行う特徴抽出と，特徴量から印象語への変換を行う印象メタデータ抽出の 2 段階で構成される．

本章で構成する $\mathcal{ML}$ の概略図を図 5.1 に示す．

これまで我々が提案してきた $\mathcal{ML}$ では，印象メタデータ抽出の作用素について，n 次元の特徴量を m 次元の印象メタデータに写像する $n \times m$ 行列，すなわち線型作用素を用いて行っ

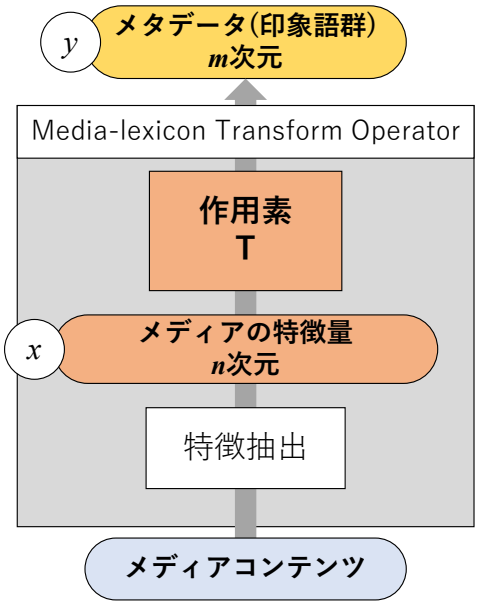


図 5.1: 印象メタデータ抽出を作用素に拡張した $\mathcal{ML}$ の概略図

てきた．本稿ではこの印象メタデータ抽出の部分を線型作用素から拡張し，線型に限らない作用素を用いて構成することとする．

特徴抽出および印象メタデータ抽出を構成することにより，各メディアコンテンツに印象メタデータを付与することが可能となる．これまでの $\mathcal{ML}$ の具体的事例においては，特徴抽出および印象メタデータ抽出を，そのメディアコンテンツの専門家の研究によって示されたメディアコンテンツと印象語の関係を用いてきた．例えば，画像の色彩から受ける印象の心理調査を行ったカラーイメージスケール [55] は画像というメディアコンテンツの専門家の研究のひとつである．

この $\mathcal{ML}$ に加え，印象メタデータ間での類似度を計量する枠組みを導入することで，異種のメディアコンテンツから抽出された印象を比較して類似度を計量することができる．例えば楽曲を入力として印象が類似している画像を検索することが可能となる．他にも，印象語群を入力して様々な種類のメディアコンテンツを検索するなど，異種メディアコンテンツを横断した検索システム [5, 6, 26] を実現することが可能となる．

また，ここで説明された $\mathcal{ML}$ はメディアコンテンツを印象メタデータへ変換する機構であるが，一方で，印象メタデータからメディアコンテンツへ変換する機構を考えることができる．これは，$\mathcal{ML}$ の逆演算に相当する．これについて，我々は $\mathcal{ML}$ の逆演算を行う統計的一般化逆作用素（$i\mathcal{ML}$）[6] を提案している．$i\mathcal{ML}$ は，$\mathcal{ML}$ の逆演算であるため，$\mathcal{ML}$ と $i\mathcal{ML}$ を組み合わせることにより，入力されたメディアデータと $i\mathcal{ML}$ が出力したメディアデータの比較が可能となる．すなわち $i\mathcal{ML}$ は，$\mathcal{ML}$ が正しく構成されたかを検証する検証系として

表 5.1: CM 好感要因 15 項目

アンケート選択肢 (好感要因)
出演者・キャラクター
ユーモラスな所が
セクシーだから
宣伝文句が印象的
音楽・サウンドが印象的
商品にひかれた
説得力に共感した
ダサイけど憎めない
時代の先端を感じた
心がなごむ
ストーリー展開がおもしろい
企業姿勢にウソがない
映像・画像がよい
周囲の評判もよい
かわいらしい

の役割を果たすといえる.

5.4 使用するテレビ CM データ

本節では，$\mathcal{ML}$ の構成に用いるテレビ CM の大規模モニター調査データと，各テレビ CM の書き起こしテキストデータが含まれた CM 映像・表現データについて述べる.

5.4.1 モニター調査データ

この調査は，株式会社東京企画 CM 総合研究所が行うもので，毎月，関東地方 1 都 6 県に住む，年齢や性別によって分かれた 15 の階層（例えば，20 代男性，18〜24 歳の独身女性，40 代主婦など）から各 200 名ずつ抽出して集められたモニター計 3,000 人に対して行われる．モニターは毎月一定の日に，テレビを消した状態で，直近の一ヶ月以内で印象に残ったテレビ CM を純粋想起によって最大 5 つまで記述する．また，モニターは純粋想起によって記述された各 CM に関して，CM 好感要因 15 項目の中から複数回答可能として選択する（ただし，必ず 1 つ以上選択する）．好感要因の詳細は表 5.1 の通りである.

この調査を集計し，純粋想起によって記述した人数を，その CM の得票数とする．また，それぞれの好感要因について，その好感要因を選択した人数を，その CM における各好感要因の得票数とする.

5.4.2 CM 映像・表現データ

このデータは，株式会社東京企画 CM 総合研究所が作成するもので，東京キー 5 局で放映されたすべての CM の情報をテキスト化したものである．CM を製作した企業名，商品の銘柄，作品名などの情報に加え，CM の主な情景・シーン，音声・セリフ，画面上の全コピーの書き起こしテキストデータが収録されている．CM の主な情景・シーン，音声・セリフ，画面上の全コピーについては，CM 総合研究所におけるマニュアルに準じて行われており，以下の方法で記述される．

CM の主な情景・シーン

動作の内容と出演者名を用いて「何々をする誰」で記述する．人が出てこない、具体的に注目すべき存在が無い場合,「～ な映像」で記述する．

音声・セリフ

画面に音声と対応する文字が存在する場合，それに準じて音声を記述する．音声のみで使用される言葉の場合は，過去作などに準じて入力者が判断して記述する．出演者の発言は「」で括り，ナレーションの場合は通常の文章で記述する．歌詞を入力する場合，“♪” を歌詞の前に付ける．

画面上の全コピー

解読可能な文章・単語を文字通りに記述する．ただし，企業名，商品名，グレード名，作品名などが登録されているマスターに準じた表記にすることがある．

また，これらの記述は以下のプロセスを経て行われる．

1. 入力者が出演タレントと上記 3 つのテキストデータを入力する．
2. 入力者とは別人によって，誤記・変換ミスを校正する．
3. 以降，誤記が判明した時点で，随時情報を修正する．

5.5 テレビ CM を対象とした $\mathcal{ML}$ の構成

本節では，テレビ CM の大規模モニター調査データと CM 映像・表現データの書き起こしテキストデータを用いて 5.3 で示した $\mathcal{ML}$ を構成し，テレビ CM の好感要因を推定する方式を示す．

本章において，$\mathcal{ML}$ のメディアコンテンツ，特徴量，印象メタデータは，それぞれテレビ CM と字コンテ，単語，好感要因と対応する．その対応付けをまとめたものを図 5.2 に示す．これまで構成されてきた $\mathcal{ML}$ がメディアコンテンツの専門家の既存の研究を用いていたのに対し，本稿の $\mathcal{ML}$ は，実データを用いて構成する点に特徴を持つ．

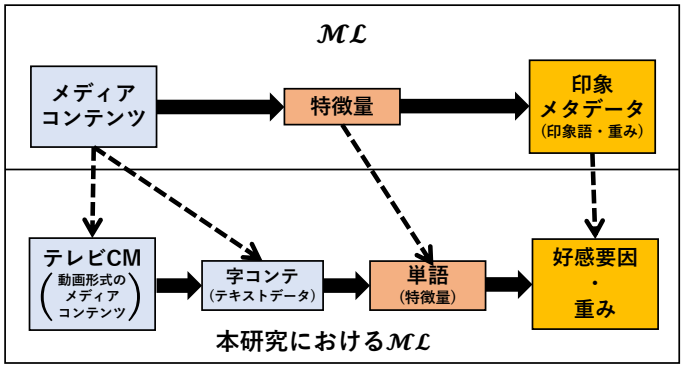


図 5.2: テレビ CM を対象とした ML における対応付け

5.5.1 テレビ CM の特徴抽出

本章では CM 映像・表現データのうち，CM の主な情景・シーン，音声・セリフ，画面上の全コピーの 3 つの書き起こしテキストデータを用いる．この 3 つの書き起こし文を連結して 1 つにまとめたものを，CM を表現するテキストデータとする．

このテキストデータに対して分かち書きを行って単語に分割したあと，Bag-of-Words[21] で特徴量を表現する．BoW は，出現する単語全てにあらかじめインデクスを割り振り，テキストデータの単語の出現回数をベクトルの要素として用いる手法である．各 CM に対して BoW を用いてテキストデータをベクトルにすることにより，行列の各行ベクトルが各 CM に対応した $s \times n$ 行列でテキストデータを表現することができる．ここで，s はシステムを構成するのに使用するテレビ CM の総数であり，n はテレビ CM の書き起こし文に存在する単語の種類の数である．

5.5.2 テレビ CM を対象とした $\mathcal{ML}$ 構成方式

本節では，テレビ CM を対象とした $\mathcal{ML}$ の構成方式の詳細について述べる．構成方式の概略図を図 5.3 に示す．本手法は図 5.3 で示すように，テレビ CM の特徴量を好感要因に変換する作用素 T の生成を行う部分と，作用素を用いたテレビ CM の好感度の推定を行う部分の 2 つに分けることができる．ここでは，実データを用いた $\mathcal{ML}$ の構成方式として，相関方式，内積方式，重回帰方式，ニューラルネットワーク方式の 4 つの方式を提案する．これらの方式について，以下に述べる．

相関方式による作用素 T の生成

ここでは，好感要因と単語の相関を表現する相関行列を生成して作用素とする，相関方式を示す．この相関行列は以下の 3 つの Step により生成される．

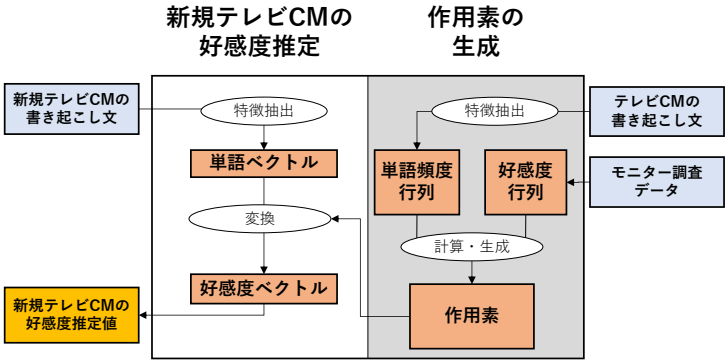


図 5.3: 提案手法の概略図

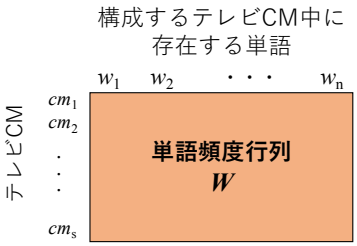


図 5.4: 単語頻度行列 $\boldsymbol{W}$ の形式

Step 1: 単語頻度行列 W の生成

図 5.4 は単語頻度行列を示したものである．この行列は，テレビ CM の書き起こし文を BoW によって行列で表現したものである．単語頻度行列は各行が各テレビ CM における単語の出現頻度を表す．ただし，今回のシステムでは出現頻度に対して，各テレビ CM をドキュメントとみなした TF-IDF による重み付けを行う．今回のシステムで使用した TF-IDF は以下の式で表されるものを用いる．

$$tfidf(cm_i, w_j, CM) = tf(cm_i, w_j) \cdot idf(w_j, CM)$$

$$tf(cm_i, w_j) = |\{w_j | w_k \in cm_i, w_k = w_j\}|$$

$$idf(w_j, CM) = log \frac{|CM|}{|\{w_j | cm_k \in CM, w_j \in cm_k\}|}$$

ただし，$|\cdot|$ は要素数を表す．

Step 2: 好感度行列 F の生成

図 5.5 は好感度行列を示したものである．この行列は 4.1 節で述べたモニター調査の結果について，対応するテレビ CM の得票数で除算し，得票割合とした上で縦に並べたものである．ま

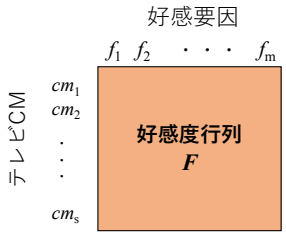


図 5.5: 好感度行列 $\boldsymbol{F}$ の形式

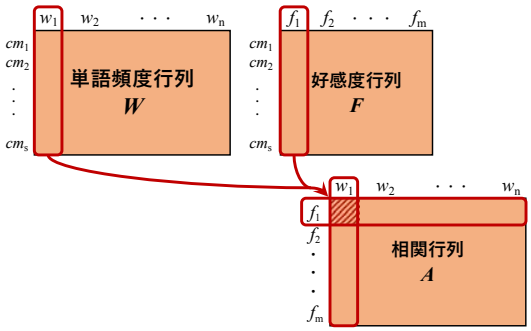


図 5.6: 相関行列の形式

た，この調査ではモニターの情報として，性別と年代を 15 の階層から選択するようになっている．この各階層ごとの得票数も上記の手法により行列化を行うことで，年代による好感度行列を作成する．

Step 3: 相関行列 $\boldsymbol{A}$ の生成

5.6 は相関行列を示したものである．この行列は前述の単語頻度行列 $\boldsymbol{W}$，好感度行列 $\boldsymbol{F}$ の関係性を表現する $m \times n$ の行列である．相関行列 $\boldsymbol{A}$ の各要素 $\boldsymbol{A}_{i,j}$ は

$$\boldsymbol{A}_{i,j} = \mathrm{cor}\left(\boldsymbol{W}_{:,i}, \boldsymbol{F}_{:,j}\right)$$

で計算される．ただし，$\mathrm{cor}\left(\boldsymbol{a}, \boldsymbol{b}\right)$ は $\boldsymbol{a}$ と $\boldsymbol{b}$ の相関係数を表し，$\boldsymbol{X}_{:,k}$ は行列 $\boldsymbol{X}$ の第 k 列を表す．相関係数を用いることにより，単語頻度行列の各行と好感度行列の各行，つまり，各単語と各好感要因の重みとの相関を数値によって表現する．要素の値が正になれば，その単語は好感要因に相関があると考えられ，負の値になれば逆相関があると考えられる．

このようにして作成された相関行列を，単語と好感要因との相関を表す変換行列の役割とし，$\mathcal{ML}$ の作用素 T とする．

内積方式による作用素 T の生成

相関係数方式では単語頻度行列の各列と好感度行列の各列の相関係数を求め，それを相関行列の各要素としたが，単語頻度行列の各列と好感度行列の各列の内積を要素とする行列についても線型作用素を考えることができる．

Step 1: 単語頻度行列 $\boldsymbol{W}$ の生成
相関係数方式と同様に行う．

Step 2: 好感度行列 $\boldsymbol{F}$ の生成
相関係数方式と同様に行う．

Step 3: 内積を要素とする行列 B の生成

Step3 における相関行列 $\boldsymbol{A}$ の代わりに内積を要素とする行列 $\boldsymbol{B}$,

$$\boldsymbol{B} = \boldsymbol{F}^{\mathsf{T}}\boldsymbol{W}$$

を用いる．

重回帰方式による作用素 T の生成

相関係数方式と内積方式では行列を $\mathcal{ML}$ の作用素として用いることによって特徴量から印象への変換を行っていたが，重回帰分析によって特徴量を印象に変換する手法も考えられる．

Step 1: 単語頻度行列 $\boldsymbol{W}$ の生成
相関係数方式と同様に行う．

Step 2: 好感度行列 $\boldsymbol{F}$ の生成
相関係数方式と同様に行う．

Step 3: 作用素の生成
CM に出現する単語 w_k の出現頻度を x_k とするとき，$i = 1, 2, \ldots, m$ に対してそれぞれの好感要因 f_i が

$$f_i = a_{i,0} + a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n + \epsilon_i$$

という式で表されると仮定する．ここで $a_{i,0}, a_{i,1}, a_{i,2}, \ldots, a_{i,n}$ は偏回帰係数であり，ϵ_i は誤差である．今回のシステムでは，単語頻度行列 $\boldsymbol{W}$ と好感度行列 $\boldsymbol{F}$ を訓練データとし，最小二

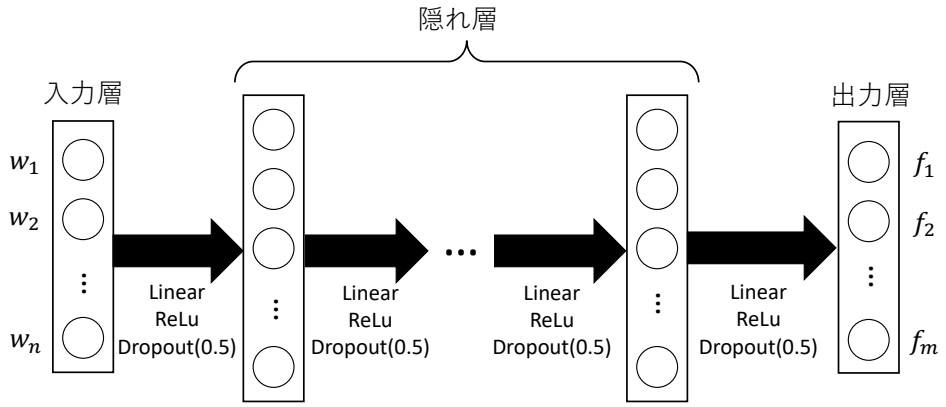


図 5.7: ニューラルネットワークの構成

乗法に L2 ノルムの正則化項を加えたリッジ回帰によって偏回帰係数 $a_{i,0}, a_{i,1}, a_{i,2}, \ldots, a_{i,n}$ を求める．この偏回帰係数を用いて，好感要因の予測値 $\hat{f}_i$ を

$$\hat{f}_i = a_{i,0} + a_{i,1}x_1 + a_{i,2}x_2 + \ldots + a_{i,n}x_n$$

で求める．この予測を行う式を作用素として用いる．

ニューラルネットワーク方式による作用素 T の生成

特徴量から印象を推定する手法としては，ニューラルネットワークを用いた手法を考えることもできる．単語頻度行列 $\boldsymbol{W}$ を訓練データ，好感度行列 $\boldsymbol{F}$ をその訓練データの正解データとして与え，ニューラルネットワークに学習させることによって得票数に対する好感要因別の得票割合を推定する．

Step 1: 単語頻度行列 $\boldsymbol{W}$ の生成
相関係数方式と同様に行う．

Step 2: 好感度行列 $\boldsymbol{F}$ の生成
相関係数方式と同様に行う．

Step 3: ニューラルネットワークの学習

ニューラルネットワークの構成を図 5.7 に示す．このニューラルネットワークは入力層と出力層，全てユニット数が等しい数の隠れ層から構成されている．入力層と出力層のユニット数は，それぞれ単語頻度行列の単語数 n および好感度行列の好感要因数 m に対応する．隠れ層のユニット数が等しいのはネットワークの構成を簡単に決定するためであり，隠れ層の数

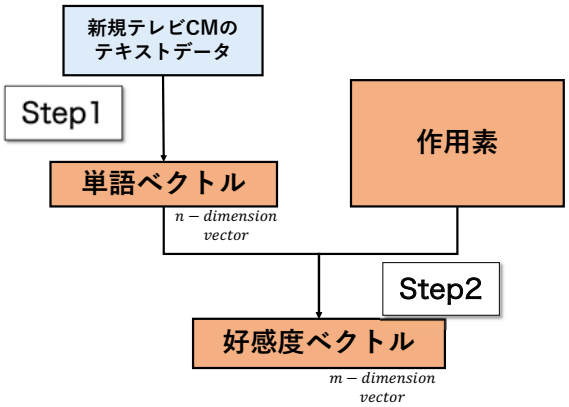


図 5.8: 新規テレビ CM の好感度推定手順

とユニット数はグリッドサーチによって決定する．それぞれの層は全結合されており，活性化関数として ReLu を用いている．また，全ての層において 0.5 の確率でドロップアウトが行われる．損失関数には平均二乗誤差 (Mean Squared Error, MSE) を用いる．勾配降下法における最適化関数として Adam を用いる．学習は 1,000 エポックまで行われるか，MSE の減少が直近 3 エポックにおいて観測できなくなった段階で終了する．以上の構成を持つニューラルネットワークに単語頻度行列の各行を訓練データ，そのデータに対応する好感度行列の各行を正解データとして与え，学習を行う．この学習させたニューラルネットワークを作用素として用いる．

5.5.3 新規テレビ CM の好感度推定

ここでは，生成した作用素を用いて，新規テレビ CM の好感度を推定する方法について述べる．推定方式の手順を図 5.8 に示す．5.8 が示すように，テレビ CM の印象の推定は以下の 2 つの Step により行われる．

Step 1: 単語ベクトル x の生成

単語ベクトル $\boldsymbol{x}$ とは，好感度推定を行う新規テレビ CM の特徴量を表すベクトルである．このベクトルは 5.5.1 で示した BoW による特徴量抽出を行うことで生成される．本来，BoW による特徴量抽出では，そのベクトルの次元数は元のテキストデータ中に存在する単語の種類の数となるが，この操作で作成する単語ベクトルは，各要素が単語頻度行列 $\boldsymbol{W}$ の各列に対応している次元数 n のベクトルとなる．この単語ベクトル $\boldsymbol{x}$ は 2 ノルムで正規化する。

Step 2: 好感度ベクトル y の出力

単語ベクトル $\boldsymbol{x}$ を各作用素によって変換することで好感要因別の重みを表す好感度ベクトル $\boldsymbol{y}$ を計算する．

5.5.2 の相関行列，5.5.2 の内積を要素とする行列による線型作用素では，

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

となる．ここで好感度ベクトル $\boldsymbol{y}$ は m 次元のベクトルであり，各要素がそれぞれの好感要因に対応する重みを表す．

5.5.2 の重回帰による手法では

$$\boldsymbol{y} = (\hat{f_1}, \hat{f_2}, \ldots, \hat{f_m})^{\mathsf{T}}$$

である．5.5.2 のニューラルネットワークによる手法では，単語ベクトルをネットワークに入力したときの，出力層の各ユニットの値を並べたものが $\boldsymbol{y}$ となる．

この操作により，テレビ CM の書き起こし文を BoW によって特徴量抽出された単語ベクトルから，そのテレビ CM の推定される好感要因とその重みを計算することが可能となる．

5.6 実験

本節では，4 節で述べた手法について，その有効性の検証を行う．

実験 1 として，実際にテレビ CM 製作を想定して作成された字コンテに対して好感度推定を行った例を示す．また，実験 2 では，実装したシステムを用いて好感度推定を行い，実際の大規模モニター調査結果を正解として本方式の検証を行う．

5.6.1 実験システム及び実験データ

テレビ CM に関するテキストデータの形態素解析及び，分かち書きには日本語向け形態素解析器である MeCab[56] を利用した．辞書には IPA 辞書に加え，流行の言葉，人名も抽出させるため，Wikipedia 日本語版に存在する見出し語のうち 3 文字以上の単語を追加登録して利用した．実験システムは Python 言語を用いて実装した．本システムは，Web ブラウザで動作する．

図 5.9 に本システムの UI を示す．左上に文章の入力フォームがあり，好感度推定を行いたい文章を入れると，ページ下部に各好感要因の推定値が棒グラフの形式で出力される．

実験には，2015 年 10 月から 2016 年 11 月に関東地方で放送された 16,358 件のテレビ CM の書き起こし文及び，それらテレビ CM に対する大規模モニター調査の結果を利用する．なお，テレビ CM の中には，出演者が登場しないものや，画面上に文字が表示されないものもある．そのため，書き起こし文にする際に，一部のテレビ CM で欠損が生じる．そこで，書き起こし文が欠損となるテレビ CM は除外して実験を行う．また，今回用いた大規模モニター調査では，モニターに 5 つまでのテレビ CM を自然想起によって回答してもらう形式である．そのため，放映されたすべてのテレビ CM の好感度を調査できておらず，得票できなかったテレビ CM では，その好感要因は不明である．また，得票数が少ないものは好感要因の推測結果に大きな偏りを起こすおそれがある．そこで，得票数が 5 未満であるテレビ CM を除外

図 5.9: Web ブラウザで動作する提案システムの UI

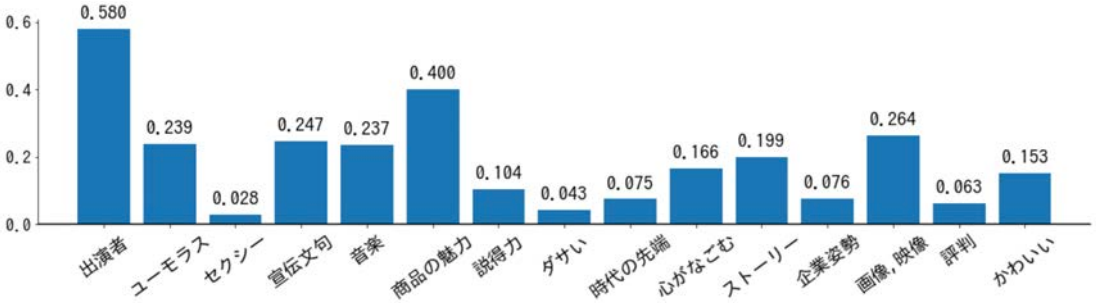


図 5.10: 訓練データにおける得票割合の平均値

する．16,358 件の全データに対して欠損によって除外された CM と得票数が 5 未満の CM を除外したのち，実験に使えるデータとして残った 3,407 件のテレビ CM を用いて実験を行う．

このうち，2,044 件を訓練データとし，行列および重回帰分析・ニューラルネットワークの構成に使用する．さらに残りの 1,361 件のうち 1,022 件をニューラルネットワークのバリデーションに使用し，残りの 341 件を最終的な一致率を検証するデータとして用いる．

単語頻度行列を生成する際に利用する単語として，書き起こし文にある単語のうち，MeCab が助詞，助動詞及び，記号と判定した単語を除いたすべての単語を用いており，動詞などの活用の存在する品詞は基本形に直して使用する．

ニューラルネットワークのユニット数，隠れ層については，隠れ数を $1, 2, 3, \ldots, 10$，ユニット数を 32,64,128, . . ., 4,096 までそれぞれ変化させ，バリデーションにおいて最も MSE の低かった隠れ層 1，ユニット数 4,096 を用いた．

5.10 に，今回の訓練データにおける得票割合の平均値を示す．この図が示すように，好感要因によって得票数が多い要因（“出演者”，“商品の魅力”，“画像，映像” など），得票数が少ない要因（“セクシー”，“ダサい”，“評判” など）があることを留意しておくべきである．

表 5.2: 実験 1：入力字コンテ

case 1
桜の花びらが舞う丘。幸せそうに家族を見つめる。 帰り道は流れ星が舞う。思い出ひらひら。
case 2
ビタミンレモンサイダーの元となるレモンの産地を 訪れる女性。 「うわぁ、すごい量のレモン！」 レモン一つをもぎ取り、丁寧に洗い、 その場で搾ってみる。 「この柑橘系の香りが、 涼しげな気分にさせてくれますよね」 農家の方がレモンの効能を説明。 「レモンを飲むと肌に潤いが出るから、化粧水いらずよ」 ペットボトルを女性に手渡し、それを飲む女性。 自然がもたらす神秘の配合 ビタミンレモンサイダー ちゃっかりレモンを持ち帰る女性。

5.6.2 実験 1（CM 好感度推定）

実験方法

訓練データを用いて CM 好感度を推定する作用素を構成して実験システムを構築する．このシステムに実際の使用を想定した字コンテとなるテキストデータを入力し，好感要因とその重みが出力されることを確認する．

表 5.2 に，好感度推定を行う 2 種類の字コンテを示す．これらの字コンテは，協力企業がテレビ CM 製作を想定して作成したものであり，本実験は，提案方式の実活用シーンと捉えることが出来る．

実験結果

システムによる好感度推定の結果を図 5.11 から図 5.18 に示す．図 5.11 に示すような，case1 で相関係数を用いた結果では，好感要因として “心がなごむ” の重みが他と比べ突出した値を示している．そのため，この字コンテをもとにして製作するテレビ CM が “心がなごむ” を視聴者へのメッセージとしている場合，製作の意図に合致していることを確認することができる．一方で，“ユーモラス” や “商品の魅力” の重みは負に大きく示していることから，訓練データの各 CM に比べると，それらの好感要因に関するメッセージ性は低いと判断すること

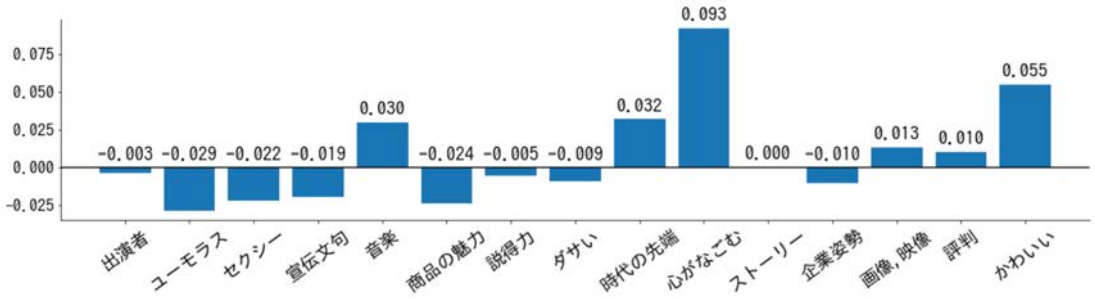


図 5.11: 実験 1：好感度予測推定結果 case1（相関方式）

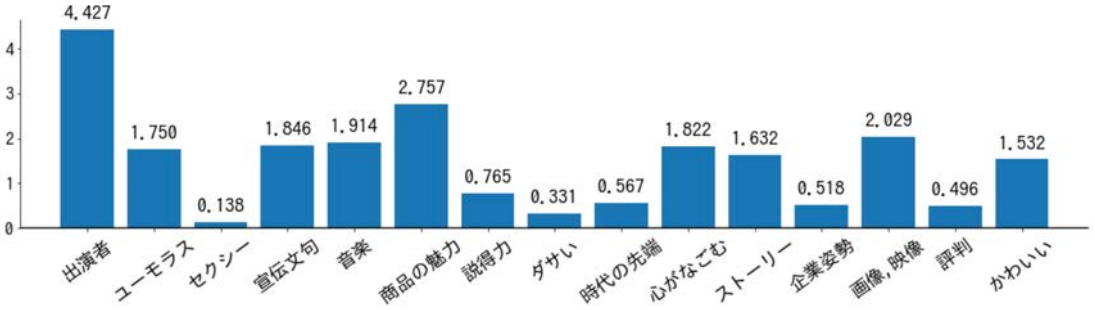


図 5.12: 実験 1：好感度予測推定結果 case1（内積方式）

ができる．この字コンテが “商品の魅力” を強く伝えたい CM であったとすれば，この結果から，字コンテを変更する必要があると推測できる．

また，相関係数以外の手法では出力される結果の傾向が似通っており，どの手法でも “出演者” が一番高い．この字コンテを CM にするにあたって，出演者の選定が重要であるという示唆であると推測できる．同様に “音楽” の選曲が重要であると判断することができる．視聴者に与える印象としては，“商品の魅力”，“心がなごむ”，“画像，映像”，“かわいい” などが高い値を示しており，これらが伝えたいメッセージであるか吟味することができる．

図 5.15 に示すような，case2 で相関係数を用いた結果では，他の CM よりも “セクシー”，“

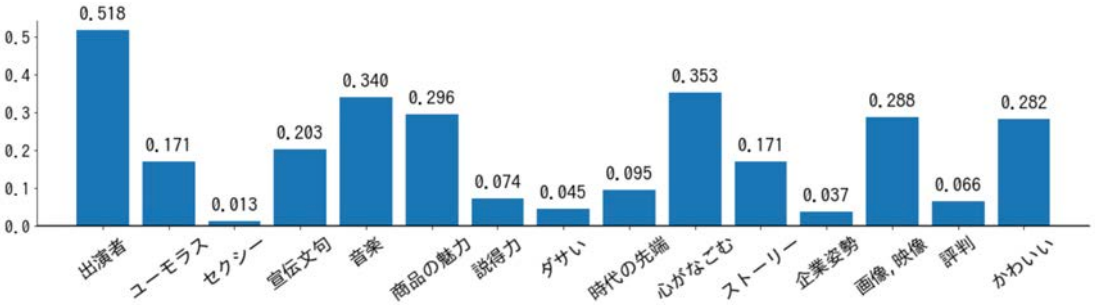


図 5.13: 実験 1：好感度予測推定結果 case1（重回帰方式）

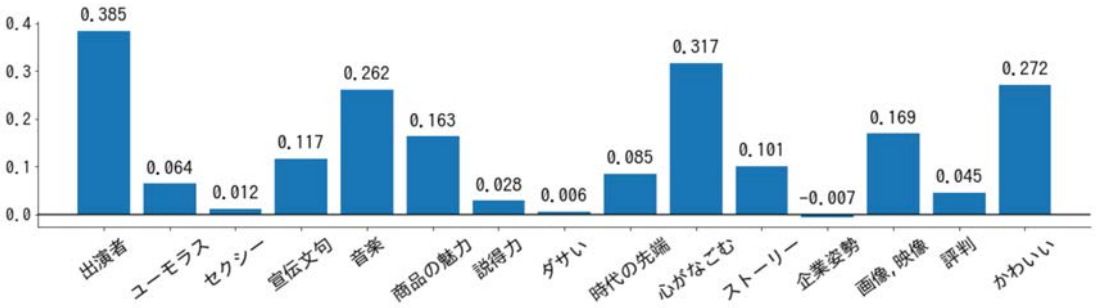


図 5.14: 実験 1：好感度予測推定結果 case1（NN 方式）

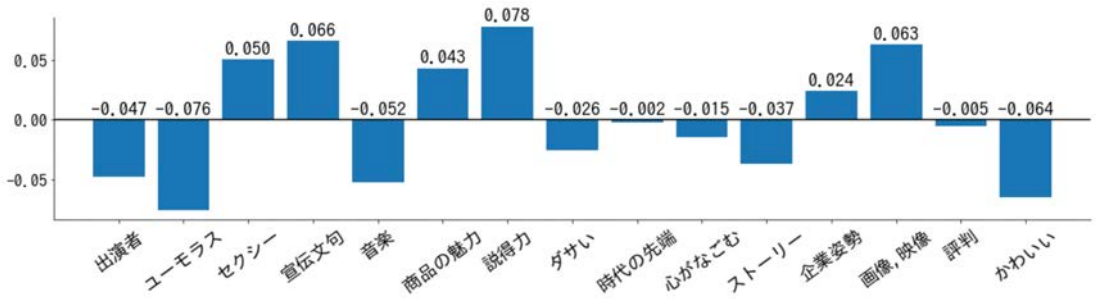


図 5.15: 実験 1：好感度推定結果結果 case2（相関方式）

宣伝文句”，“商品の魅力”，“説得力”，“画像，映像” を伝えやすい CM であると判断できる．

また，相関係数以外の手法では “出演者” の他に “宣伝文句”，“商品の魅力”，“画像，映像” が高いことが示されている．

このように，相関係数を用いる $\mathcal{ML}$ ではその式の性質から，訓練データとして与えた CM と比べた場合の好感要因の傾向を示す．内積を用いる $\mathcal{ML}$ ではその CM での好感要因の比較が可能である．また，重回帰・NN を用いる手法では，その CM での好感要因の比較に加え，実際に得票数の割合を推測することが可能である．

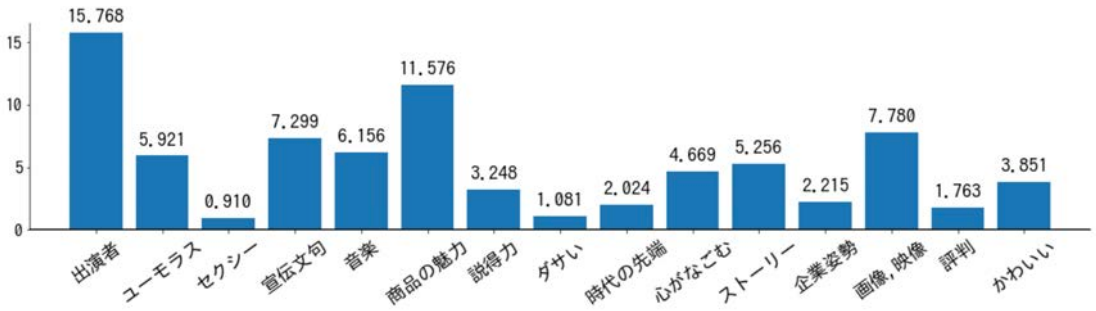


図 5.16: 実験 1：好感度予測推定結果 case2（内積方式）

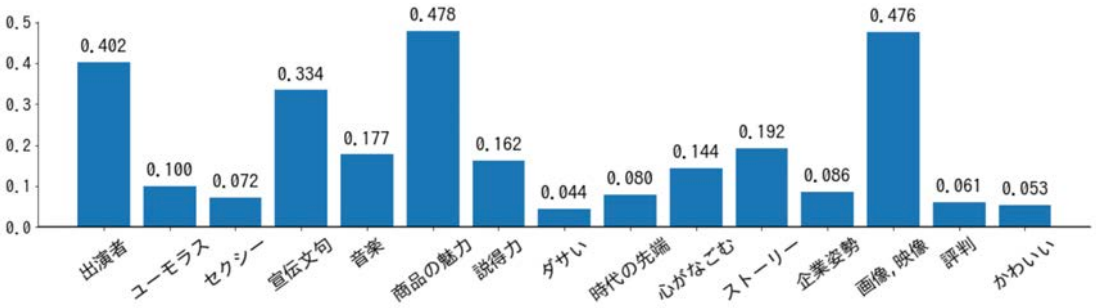


図 5.17: 実験 1：好感度予測推定結果 case2（重回帰方式）

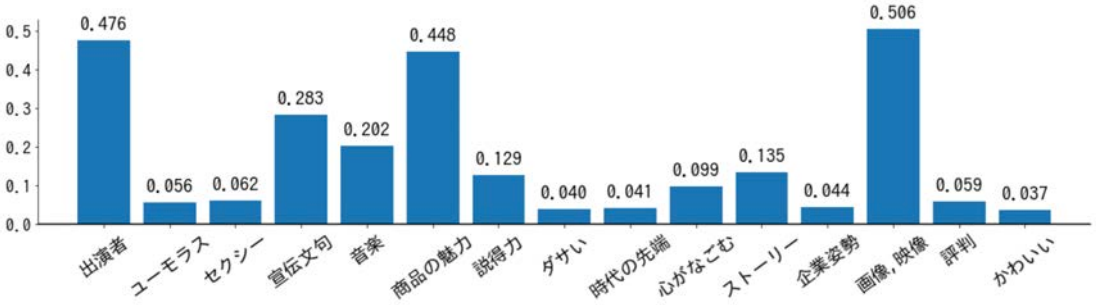


図 5.18: 実験 1：好感度予測推定結果 case2（NN 方式）

5.6.3 実験 2（推定性能調査）

実験方法

相関方式による $\mathcal{ML}$，内積方式による $\mathcal{ML}$，重回帰方式による $\mathcal{ML}$，ニューラルネットワーク方式による $\mathcal{ML}$ を用いた好感度推定方式について，評価指標として順位内一致率を導入し，それぞれの方式の有効性を検証する．これらの方式に加え，ベースラインとして，好感要因の順位を無作為に決定する方式での好感要因の推定を行う．この無作為抽出は，大規模モニター調査の結果から最尤推定によって求めた各好感要因の得票の確率を用いて行う．この実験では，訓練データから生成した作用素を用いてテストデータを好感要因に変換し，テストデータの実際の値と比較することで評価を行う．しかし，テストデータは，その CM を想起したモニターの数に対する好感要因の得票数の割合であり，必ず 0 以上の値をとる．今回生成した作用素のうち相関係数を用いた $\mathcal{ML}$ による推定では，推定結果は負の値も取りうる．また，内積を用いた $\mathcal{ML}$ による推定では得票割合の最大値を推定することができないため，実際の値に合わせた正規化ができない．すなわち，この 2 つの方式においては，実際の値と推定された値の差には直接の意味がない．そこで，全ての手法を統一的に評価するために，実際のデータ・推定値のどちらに関しても，数値が正に大きいものほど製作する CM に強く影響する好感要因であると考え，好感要因の順位の一致率を見てシステムを検証する．これは，今回のシステムを製作するにあたって，協力企業から「実際の値ではなく，その CM の

上位数件の好感要因が分かれば良い」という示唆を頂いたことによる．具体的には，好感要因上位 r 件までのランキングを実際の値と推定値それぞれに関して $r = 1, 2, \ldots, m$ まで作成し，各ランキングの好感要因の一致率によって評価を行う．好感要因上位 r 件の一致率 $f(r)$ を以下のように定義する．

$$f(r) = \frac{\left|F_E^r \cap F_Q^r\right|}{r}$$

F_E^r は，好感要因の推定によって得られた好感要因の各値に順位付けを行い，上位 r 位までの重みを持つ好感要因による集合であり，F_Q^r は，アンケート調査によって得られた好感要因への得票数に順位付けを行い，上位 r 位までの好感要因による集合である．なお，推定値が同じ，または得票数が同じ好感要因が存在する場合，それらは順位が低いほうの同率順位として扱うこととする．例えば 3 位と 4 位が同じ値である場合，どちらも 4 位とする．これは，一致率 $f(r)$ が 1 を越えないようにするための措置である．

実験結果

図 5.19 に，順位づけによる評価を行った結果を示す．縦軸は該当順位までの評価値，横軸に順位を表している．順位が 15 の時，全ての好感要因が評価の対象となるため，評価値は必ず 1 になる．

まずベースラインとなる最尤推定によるランダムな推定については，相関係数の 9 位以降で逆転が見られるものの，上位 8 件まではどの手法よりも低い一致率を示す．この結果から，単純な得票数の割合を用いるよりも，テキストデータ内の単語に着目した手法のほうが良い結果を出すことがわかる．

手法別に見てみると，ニューラルネットワークが一番良く，重回帰がそれに拮抗している．内積と相関係数はそれら 2 つに大きく差を付ける結果になっている．これは，ニューラルネットワークや重回帰が，得票割合の推定値を出力していることが大きく影響していると考えられる．ニューラルネットワークと重回帰について，ニューラルネットワークは単語の共起性を考慮していて，重回帰では単語の共起性を考慮していないという違いがあるにもかかわらず大きな差にならなかったのは，ニューラルネットワークが単語の共起性から特徴を見つけ出せなかったと考えられる．今回の単語頻度行列の単語数 n が 15,625 語であるのに対し，訓練データとして与えられた CM の平均単語出現数は 40.71 語であり，今回の単語頻度行列は 99.74% が 0 である疎行列であったことが理由として挙げられる．また，CM そのものの性質として，オリジナル性を高めるため，他の CM とは違うキャッチコピーやセリフを使う傾向にあり，複数の CM に同じような単語の共起が多く発生しなかったことも理由として挙げられる．内積と相関係数の差については，相関係数が平均との差を基準に好感要因を判定していることが大きく影響したと考えられる．

今回の結果ではニューラルネットワークが最も良い結果を示したが，これらの手法は目的に応じて適宜選択されるべきである．字コンテに応じた得票割合を正確に求めたい場合はこの結果が示すようにニューラルネットワークで良い．ここから派生して，視聴者に伝えたい

表 5.3: 実験 2：手法別の可・不可

手法名	得票割合の推定	順位の一致率	逆演算
相関係数	×	×	○
内積	×	△	○
重回帰	○	○	△
NN	○	○	×

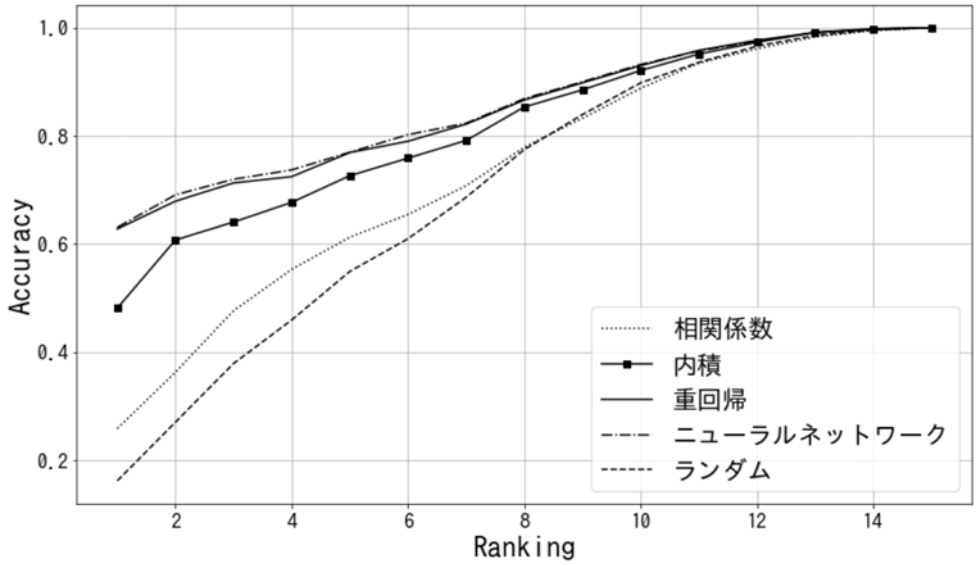


図 5.19: 実験 2：ランキングによる好感度推定の評価

好感要因を入力とし，使用すべき単語を出力するシステム，すなわち今回の $\mathcal{ML}$ の逆演算に相当する $i\mathcal{ML}$ を構成するのであれば，線型作用素である相関行列，または内積を要素とする行列のほうが都合が良い．行列の一般化逆行列を求めることにより，$i\mathcal{ML}$ を構成できるからである．重回帰では，$f_1, f_2, \ldots, f_m$ の偏回帰係数を要素とする行列についての逆行列を考えることができるが，逆行列によって出力された結果には定数項の影響があることを考慮せねばならない．ニューラルネットワークの場合では，好感要因を入力とし，単語を出力するネットワークにして学習すれば可能であるが，完全な逆演算とはならない．これらをまとめたものを表 5.3 に示す．

5.7 本章の結論

本章では，テレビ CM の好感度に関する大規模モニター調査データを用いて，テレビ CM 作品を表す単語と好感要因の関係を計算することによるテレビ CM 好感度推定方式を示した．これは，新たに製作されるテレビ CM を対象とし，それをテキストデータで表現したものか

ら，そのテレビ CM が持つ好感要因とその重みを推定し出力するものである．また，この方式を実装し，実験を行い，本方式の有効性を示した．

本方式の実現により，モニター調査の結果を活用し，過去に放送された CM を元に新しい CM の好感度の推定値を，CM の製作過程の段階であっても算出することが出来るようになる．したがって CM クリエイターは，新しく CM を製作する段階においても，製作中のテレビ CM の好感度の推定値を知ることが可能となり，テレビ CM を製作する意図と合致するかを判断することができるようになる．

本方式では，テレビ CM を表現するデータとして，テキストデータを利用し，加えてテレビ CM の好感度評価としてアンケート調査による結果を利用している．したがって，メディアコンテンツと印象を表す好感度データのセットが用意されていれば，任意のメディアコンテンツ、例えば動画コンテンツやテレビ番組に適用可能である．すなわち，本方式を応用することで，これまで専門家の研究が十分でなかった対象分野のメディアコンテンツに対しても $\mathcal{ML}$ を構築することが可能となり，メディアコンテンツを扱う様々なコンテンツマネジメントシステムに応用することができる．また，本方式は任意の単語と好感度を紐付ける手法であり，対応した言語のメディアコンテンツと印象を表す好感度データのセットがあれば，日本語に限らず応用が可能となる手法である．

今後解決すべき課題として，新規のテレビ CM に登場しうる未知語の処理方法がある．本稿で提案した方式では，考慮することのできる単語は，$\mathcal{ML}$ を構築した際に用いたテレビ CM の書き起こし文の単語のみであり，それ以外の単語は結果に反映させることができない．特にテレビ CM 製作のようにクリエイティブ性の高い分野では，未知語の出現頻度は高いと考えられる．この課題への対応方法として，テレビ CM の書き起こし文の特徴抽出の際に使用した BoW を，Word2Vec[39, 40, 41] などの単語の分散表現方式で置き換える方法が検討できる．また他に挙げられる課題としては，他のメディアコンテンツに対するモニター調査の結果を用いた $\mathcal{ML}$ の構成方式や，本方式の逆演算に相当する，Stochastic Generalized inversed Media-lexicon Transformation Operator($i\mathcal{ML}$) の構成方式の考案がある．

第6章　Webアクセスログからのデモグラフィックデータ予測

6.1　研究の背景

今日，インターネットを通して生成されるデータの量は急速に増加している．インターネットのユーザがコンピュータやモバイルデバイスで情報収集などの活動をする傍ら，多くのユーザの無意識のうちに，その活動に関するデータが生成され続け，蓄積・収集されている．そのデータのひとつとして挙げられるのがWebの訪問履歴であるアクセスログである．アクセスログはユーザが接触したWeb上コンテンツとそのコンテンツ間の遷移を示すものであり，ユーザの趣味嗜好や属性を反映したものである．このアクセスログを分析することはユーザのWeb上の活動を理解するために重要である．

また一方で，アクセスログには様々なWebページが記録されているため，Webページの特徴を分析することも重要なことである．Webページ上には様々な形式のコンテンツが存在し，例えばテキスト，画像，映像といったものから，テキストの色使いや言葉遣い，ページ構成などもコンテンツの一要素である．これらのコンテンツはそのページを閲覧したユーザの趣味嗜好や属性と関連があると考えられる．このようなWebページのコンテンツとユーザの属性の間の関係を明らかにすることによって，ユーザの属性に合わせたコンテンツの生成や，その逆に，接触したコンテンツからユーザの属性を推測するといったことが可能となる．

このようなデータの収集を行う目的のひとつは，企業における商品の販売の対象となる顧客グループを設定するターゲティングを行うためである．これはマイクロマーケティングと呼ばれ，顧客関係管理 (CRM, Customer Relationship Management) のプロセスのひとつとして小売業で発展してきた．顧客関係管理は既存の顧客の満足度の維持や向上を目指すとともに，新規顧客の獲得のために，既存顧客の分析を行うことを目指す．

例えばECサイトを運営しているような企業のケースでは，会員登録機能によって企業側は顧客のデモグラフィックデータを獲得することができ，そのデータを既存顧客の分析に活用することができる．しかし，潜在的な新規顧客である多くのユーザに対してはデモグラフィックデータが結び付けられていないため，このような分析を行うことができない．潜在的な新規顧客ユーザに対しても戦略的にアプローチすべきであると考えられるため，そのようなユーザが接触するWebコンテンツとユーザのデモグラフィックデータの関係を明らかにすれば，潜在的な新規顧客へのアプローチが可能となる．

今回取り扱うアクセスログは，ユーザを一意に識別できるトラッキングホストと，そのホストがアクセスしたURIと時刻を格納しているログであり，Web広告企業のプラットフォーム

上で取得されたものである．ここでの研究の目的は，この URI から得られる情報を基に，トラッキングホストのデモグラフィックデータを予測することである．さらに，その予測データを用いて，効果的な Web コンテンツを抽出することを目指す．これを実現するために，Web アクセスログを意味的に分析し，ユーザの行動データとデモグラフィックデータを相互に接続して関係を説明するモデルとして，Action-Demographic interconnection model を提案する．このモデルは 3 つの要素から成り立つ．それらは，ユーザが生成する行動データ，ユーザのデモグラフィックデータと，それらを相互接続する機能である．さらに，今回のモデルでは行動データを別の表現に変換するための外部データが行動データに接続されている．

また，提案モデルの実現のため，行動データである Web アクセスログのベクトル化手法を 4 つ提案する．Web アクセスログに格納されている URI から，コンテンツを取り出し，それを元にベクトルを構成する．今回対象とするコンテンツはテキストと画像である．これらを用いて意味的に Web アクセスログを分析し，デモグラフィックデータを予測するモデルを提案する．

6.2 関連研究

6.2.1 分類に関するアルゴリズム

分類とは，ある事象を認識，区別，識別，理解するためのプロセスである [57]．マーケティングにおいては，顧客のターゲティングのために，顧客をいくつかのカテゴリーに分類することが行われる．このカテゴリーは年齢・性別といったデモグラフィックデータに応じて行われることもあれば，趣味嗜好や，興味などで分類することもある．企業はユーザのカテゴリーに応じて適切な戦略を設定し，既存顧客の維持や新たな顧客獲得を行う．

コンピュータサイエンスの分野においては，データの集合をアルゴリズムによっていくつかのまとまりに分割する手法に関して，分割したまとまりに付与するラベルがあらかじめ存在する場合を分類と呼び，はじめにデータを分割してまとまりを作り後からその意味を考える場合をクラスタリングと呼ぶ．また，正解となるラベルの有無に着目して，前者を教師あり (学習)，後者を教師なし (学習) と呼ぶこともある．また，学習段階にて，正解となるラベルが一部のデータにのみ付与されており，その情報を用いながらラベルがないデータもまとめて分割する方法は半教師あり (学習) と呼ばれる．

今回の我々の研究では正解データが用意されているデータのみを取り扱うため，教師あり学習を採用している手法を用いる．Random Forest[58] はその 1 つで，分類や回帰を行うアンサンブル学習の手法として知られている．また，XGBoost(eXtreme Gradient Boosting)[59] も同様にアンサンブル学習の手法である．また，その他にも SVM(Support Vector Machines)[60], Nearest Neighbors[61], Naïve Bayes[62] などの手法がある．これらのアルゴリズムを，提案モデルにおける相互接続の機能として使用する．

表 6.1: 石垣らによる消費者のライフスタイルカテゴリーの定義

	Lifestyle category of customers
No.1	Conscious-consumption type
No.2	Fulfilling life-consumption type
No.3	Active-consumption type
No.4	Economical-consumption type
No.5	Planned-consumption type
No.6	Brief-consumption type

6.2.2 カテゴリーマイニング

石垣らは，PdLSI(probabilistic double-latent semantic indexing) model を用いて，消費者のカテゴリーマイニングを行った [63]．PdLSI は attention-based の PLSI[64] である．(ここでいう attention は，昨今の Neural Network における “注意”，“注目” とは別の概念である)．この研究では消費者を 6 つのカテゴリーに分類した．そのカテゴリーを表 6.1 に示す．この分類について，ランダム，k-means, PLSI, PdLSI の 4 つの手法で行い，平均絶対誤差 (Mean Absolute Error, MAE) で性能比較を行った．その結果，PdLSI が最も良い結果を示したとしている．

このライフスタイルカテゴリーは，消費者のデモグラフィックデータを抽象化したものと捉えることができる．我々の提案モデルは，この研究における ID-POS データが行動データとして一般化されたものである．さらに，我々の研究では，Web アクセスログに外部データを紐づけて深掘りすることにより，行動データそのものではなく，Web コンテンツとデモグラフィックデータの関係を見出す．

6.3 Action-Demographic Interconnection Model

ここでは我々の提案する Action-Demographic Interconnection Model について述べる．これはユーザの行動データとユーザのデモグラフィックデータを相互接続する機能をもつモデルである．概要図を図 6.1 に示す．このモデルは 3 つの要素で構成されている．

1 つ目の要素はユーザの行動を記録した行動データである．行動データは時系列データとして表現されることが多い．今回の研究においては，ユーザの Web ページの訪問履歴が行動データに当たる．

2 つめの要素はデモグラフィックデータである．今回の研究においてはユーザの会員登録情報や，ユーザへアンケート調査結果を元に収集された年齢や性別のデータがそれに当たる．

3 つめは行動データとデモグラフィックデータを相互接続する機能である．デモグラフィックデータを元に行動データを予測する機能と，その逆となる行動データからデモグラフィックデータを予測する機能を併せもつ．

これら 3 つの予想に加え，行動データを別の表現に変換するための外部データが接続されている．外部データは行動データに意味的なメタデータを付与する役割をもつ．

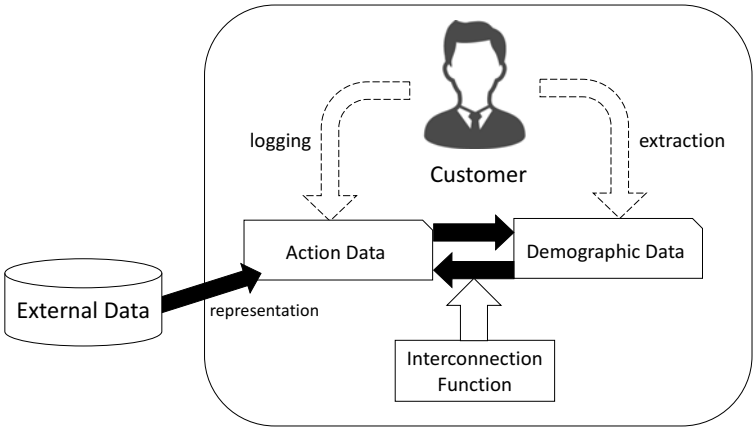


図 6.1: Action-demographic interconnection model

このモデルが重要とするのは，行動データとデモグラフィックデータの単純な関係性だけではなく，外部データによって付与されたメタデータとデモグラフィックデータの関係性を発見することにある．今回の研究ではユーザの行動データは Web アクセスログとして記録されているが，提案モデルが目指すのは，訪問した URI とデモグラフィックデータを単純に結びつけるのではなく，訪問した URI に含まれる Web コンテンツの意味的な情報とデモグラフィックデータの関係性を発見することにある．これは，デモグラフィックデータに応じたコンテンツベースの自動メディアコンテンツ生成を実現するものである．

6.4 データの統計的分析結果

今回，協力企業から Web アクセスログに関する 2 種類のデータを提供していただいた．協力企業名は NDA により非公開であるが，Web 広告プラットフォームを運営する日本の企業である．

6.4.1 History Log

1 つ目のログは，“History Log” と呼ばれるログである．今回の研究の検証用に 669,427 件のレコードを提供していただいている．その例を表 6.2 に示す．それぞれのレコードはトラッキングデータ (ホストアドレス)，訪問日付，訪問先 URI の 3 つのデータが記録されている．トラッキングデータはユニークなユーザ名を表し，日付はタイムスタンプの形式で記録されている．URI は，そのユーザがアクセスした URI である．

このレコードは以下のようなデータが格納されている．

- 49,139 のユニークなユーザ

表 6.2: History Log の例

トラッキングデータ (Customer)	日付 (yyyymmdd)	URI
hostnameA.network.com	20160106	www.webpage-a.jp/index.html
hostnameA.network.com	20160111	www.webpage-a.jp/about.html
hostnameB.japan-net.com	20160111	www.webpage-b.com/books.html
hostnameC.network.com	20160116	www.webpage-b.com/books.html
hostnameC.network.com	20160116	webpage-e.com/purchase.html
⋮	⋮	⋮

表 6.3: Tracking Data の例

トラッキングデータ (Customer)	年齢	性別 (1:男性,2:女性)
hostnameA.network.com	41	2
hostnameB.japan-net.com	40	2
hostnameC.network.com	38	1
hostnameD.fastnet.jp	24	1
hostnameE.network.com	20	2
⋮	⋮	⋮

- 134,601 のユニークな URI
- データの獲得期間は 2015 年 11 月 28 日から 2017 年 8 月 14 日の 627 日間

6.4.2 Audience Log

2 つ目のログは，“Audience Log” と呼ばれるログである．今回の検証用に切り分けられた 200,000 件が含まれている．アンケート調査を専門に行う企業 2 社によって Web 広告プラットフォームの企業に提供されており，2 社それぞれが独自に行ったアンケートの結果をマージして作成されたログである．

表 6.3 にログの例を示す．それぞれのレコードはトラッキングデータ (ホストアドレス)，年齢，性別の 3 つのカラムを持つ．アンケート調査であるため，エラーを含んでいるレコードが存在し，例えば年齢が 0 歳であったり，年齢が 100 歳以上であったりするようなレコードが含まれている．それらは，実験の際には取り除いて使用する．

図 6.2 は今回の実験データに含まれている年齢の分布をヒストグラムで表現したものである．40 歳をメインボリュームとしており，年齢ごとに均等に分けられたデータではないことに注意する．さらに，アンケート調査会社 2 社で年齢の調査方法が異なっているために，1 際刻みのデータ数にも大きく偏りがある．アンケート調査会社のうち 1 社は年齢を数値で入力

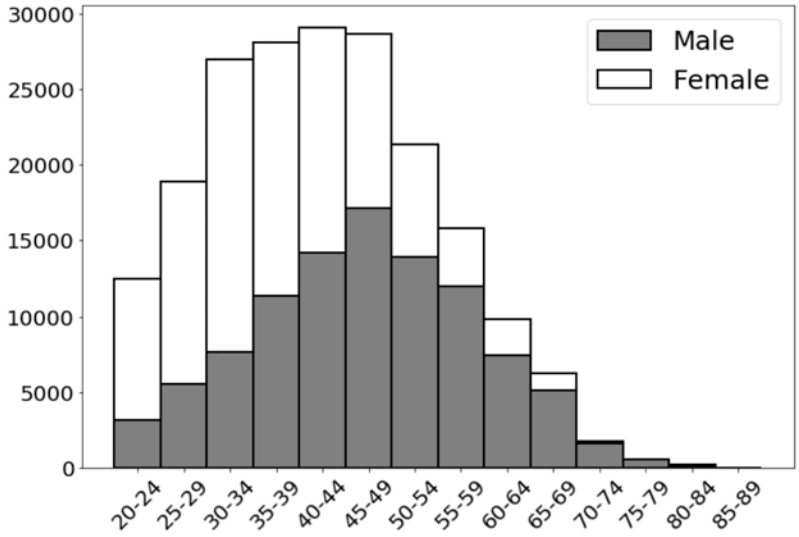


図 6.2: Audience Log に含まれる年齢のヒストグラム

させるため，1 歳刻みのデータが得られているが，もう 1 社は 5 歳刻み (20〜24 歳,25〜29 歳 . . .) の項目から選択させる方式となっており，5 歳刻みでの年齢しか取得できていない (20〜24 歳は 20，25〜29 歳は 25，. . .) といった形で格納されている)．また，今回の検証にあたってデータを切り分ける際，データ全体での男女比は 1 対 1 になるように調整されているが，年齢別に均等にする処理は行われていないため，それぞれの年齢においては男女のデータ数に差があることに注意する．

6.5 Web アクセスログのベクトル化手法

この節では，Web アクセスログの 4 つのベクトル化手法を述べる．注目する Web コンテンツによって様々な手法が考えられるが，今回はテキストと画像に注目してベクトル化を行った．

6.5.1 URI Frequency Vectors

この手法は，レコードに格納されている URI をそのまま行動データとして使用する手法である．概要図を図 6.3 に示す．ベクトルの各要素は，それぞれのユーザが訪れた URI ごとの回数である．しかし，URI をそのまま使用してしまうと非常に疎なベクトルが構成されてしまうため，ドメインレベルでのカウントとしている．History Log には 2,754 のドメインが含まれているため，この手法によって構成されるベクトルは 2,754 次元のベクトルとなる．

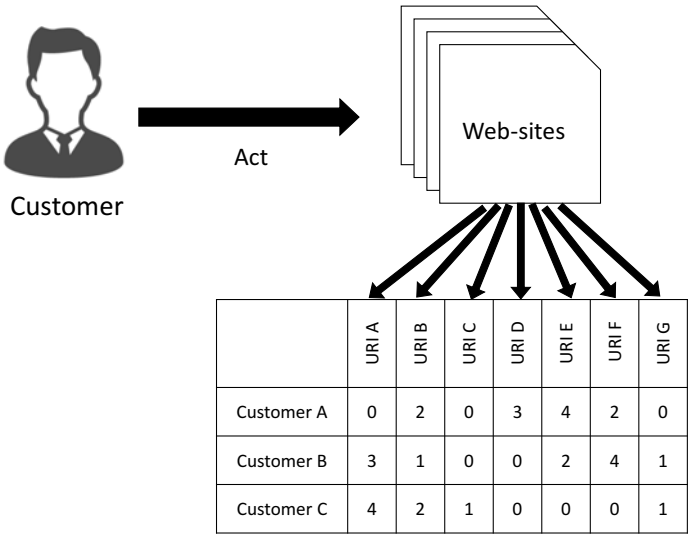


図 6.3: URI frequency vectorization

6.5.2 Word Frequency Vectors

この手法は，外部データを用いてテキストベースでベクトル化を行う手法である．テキストデータは Web コンテンツにおいて重要なコンテンツであると考えられ，意味的なコンテンツ間の関連も検証しやすい．

ここでは Bag-of-Words を用いて，ベクトル化を行う．概要図を図 6.4 に示す．今回のベクトル化においては対象となる品詞を名詞に限定し，各ユーザが訪れたページ内に存在する単語を数え上げてベクトルとする．そのため，ベクトルの次元数は，今回のデータに含まれる単語の語彙数に依存する．

6.5.3 Word Cluster Frequency Vectors

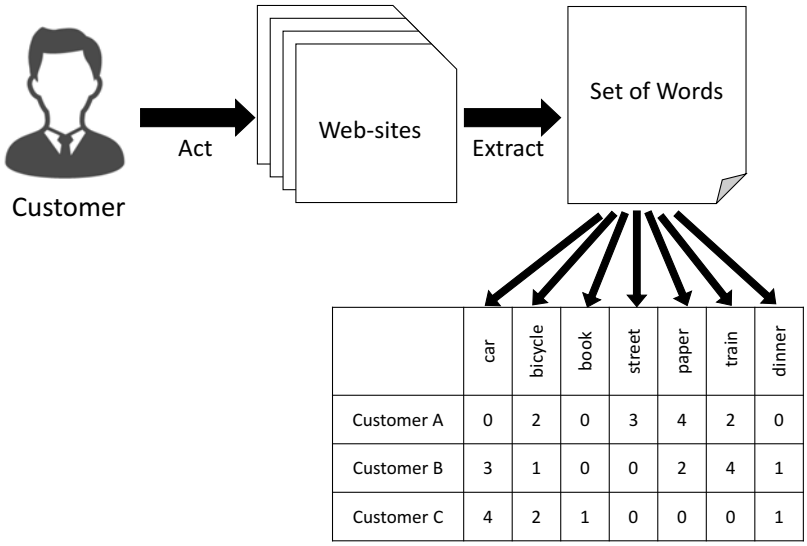
この手法は，単語間の類似度を計量することが可能な分散表現型モデルの Word2Vec[39, 40, 41] を用いる手法である．概要図を図 6.5 に示す．

前述の Word Frequency Vectors と同じように，品詞を名詞に限定する．名詞を抽出したのち，Word2Vec モデルを用いて，概念が似ていると判断される単語をまとめてクラスターを生成する．この手法では，次元数はクラスターの数に依存する．

この章における実験では，以下のようにクラスターを生成する．

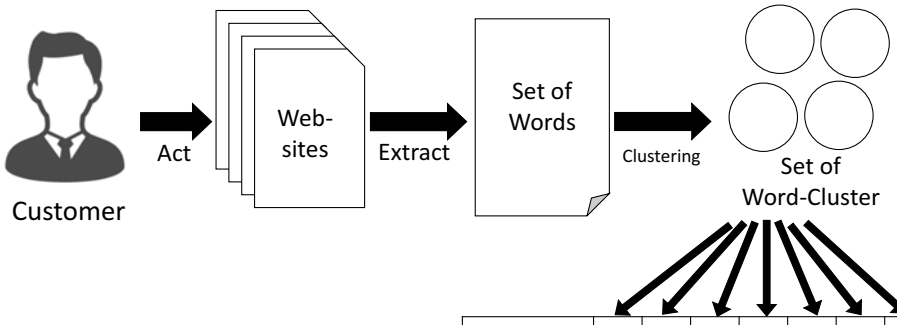
Step 1:名詞の選択

全 URI から抽出したすべての名詞からランダムに 1 つ名詞を選ぶ．



	car	bicycle	book	street	paper	train	dinner
Customer A	0	2	0	3	4	2	0
Customer B	3	1	0	0	2	4	1
Customer C	4	2	1	0	0	0	1

図 6.4: Word frequency vectorization



	Word-Cluster A	Word-Cluster B	Word-Cluster C	Word-Cluster D	Word-Cluster E	Word-Cluster F	Word-Cluster G
Customer A	0	7	2	1	2	8	5
Customer B	3	6	4	2	2	4	5
Customer C	4	2	1	7	3	4	6

図 6.5: Word cluster frequency vectorization

Step 2:クラスターへの所属

選んだ名詞と，既に生成されているクラスターに近い概念だった場合，選んだ名詞はそのクラスターに所属させる．概念の近さは，Word2Vec における類似度で判断し，閾値を定めて所属させるか決定する．

どの生成済みクラスターとも遠い概念であると判断された場合，選んだ名詞を代表とする新たなクラスターを生成する．

Step 3:全名詞に対する繰り返し

Step 1 に戻り，すべての名詞がクラスターに所属するまで繰り返し行う．

ここでは上記のような手法を使用しているが，一般的に使用されるクラスタリングの手法 (例えば k-means など) を適用することも可能である．

6.5.4 Image Histogram Vectors

4 つ目の手法は，Web ページのスクリーンショットを撮影し，画像のヒストグラムを構成する手法である．概要図を図 6.6 に示す．

History Log データに含まれている 2,754 ページに対しアクセスを行い，そのページ最上部のスクリーンショットを撮影する．撮影したスクリーンショットを RGB の 3 色に分割し，それぞれのカラーにおいて，0〜255 の画素値をカウントし，256 個の値を持つヒストグラムを構成する．最後にこのヒストグラムをベクトルとして捉え，RGB の順に要素を並べた 768 次元のベクトルを作り，そのページを表現するベクトルとする．複数のページを訪れているユーザの場合は，各要素を足し合わせて平均を取ったベクトルとした．

6.6 提案モデルの具体的な実装方法

この節では Action-Demographic Interconnection Model における Interconnection Function の実装方法を述べる．この相互接続機能は，順方向への演算を行う作用素とその逆演算を行う作用素のペアで構成される．なおここでの順方向と逆方向については便宜的なものであり，どちらが順方向であるかについては各自の定義に委ねられる．

6.6.1 順演算の作用素

本章においては，行動データから年齢と性別を予測する方向についてを順方向と定める．この作用素は，分類や回帰を行うアルゴリズムによって実現が可能である．後述の実験においては，Random Forest，XGBoost，ニューラルネットワークの 3 種類を採用して実験を行っている．またこれら 3 種類はすべて分類・回帰のどちらも行うことができるため，回帰問題として解く年齢予測，分類問題として解く性別予測について，3 種類の方法を適用して実験を行っている．

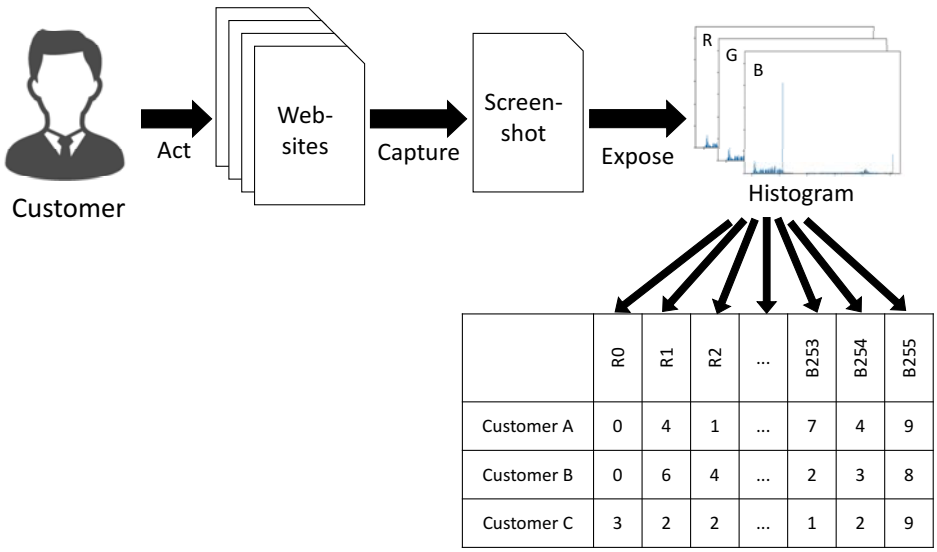


図 6.6: Image histogram vectorization

6.6.2 逆演算の作用素

上記の順方向とは逆に，年齢や性別データから Web コンテンツの特徴ベクトルを生成する部分となる．ベクトルの構成については分類や回帰といった手法で実現できるが，そこから Web コンテンツそのものを自動構成することは不良設定問題となってしまい複雑な構成となるため，今回は逆方向については取り扱わない．また，順方向の作用素を線形作用素で構成するなどの条件はあるものの，逆演算の作用素を独自に定義せず，順方向で用いた作用素の逆演算を作り出す方法もある．

6.7 デモグラフィックデータ予測の実験

この節では，Action-Demographic Interconnection Function の順方向にあたる行動データからのデモグラフィックデータ予測を行い，検証を行う．

6.7.1 研究の目的

今回の検証では，Action-Demographic Interconnection Model が実際に構築できることを示す．また，いくつかのアルゴリズムを使用して構築し，モデルの性能を比較する．今回採用するアルゴリズムは，Random Forest, XGBoost，ニューラルネットワークの 3 種類である．

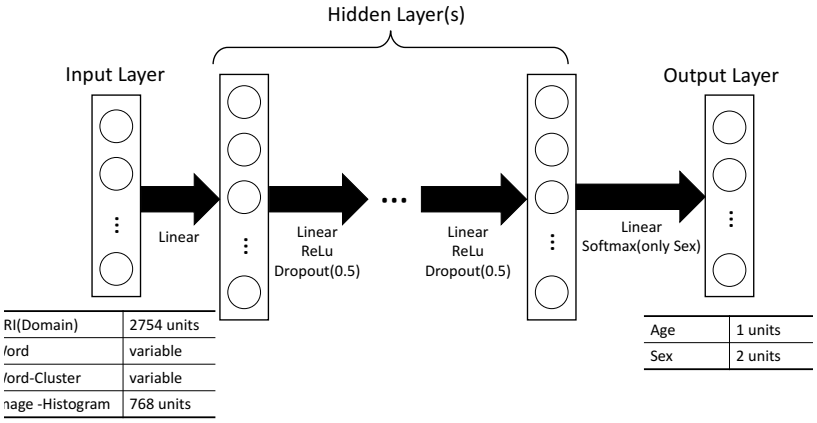


図 6.7: ニューラルネットワークの構成

6.7.2 実験環境と条件

今回の実験に関して，すべて Python でモデルの構築を行っている．Random Forest は，Topic modelling library である *gensim*[1]，XGBoost は *XGBoost*[2]，ニューラルネットワークはディープラーニングのフレームワークである *Keras*[3] とそのバックエンドとして *TensorFlow*[4] を使用した．

評価指標として正解率を用いるが，年齢を予測する場合においては 5 歳以下の誤差まで正解と認めることとした．この誤差については，協力企業からのアドバイスによる．また，すべての実験について $k = 4$ で k 分割交差検証を行っており，以下に示す結果は 4 回検証を行った平均値である．

6.7.3 ニューラルネットワークの構成

ニューラルネットワークの構成を図 6.7 に示す．

すべての層は全結合である．また，すべての中間層において活性化関数として ReLU を用いている．同様に，すべての中間層に $rate = 0.5$ となる Dropout を適用している．入力層のユニット数は，採用するベクトル化手法に依存する．出力層のユニット数は予測する対象によって異なり，年齢の場合は，年齢を直接予測するユニットが 1 つ，性別の場合は，男性である確率と女性である確率をそれぞれ予測するユニットが 1 つずつ，計 2 つとなる．中間層のユニット数については，大まかなグリッドサーチによる予備実験によって数を決定をする．この詳細は付録を参照のこと．

[1] https://radimrehurek.com/gensim/
[2] http://xgboost.readthedocs.io/en/latest/
[3] https://keras.io/
[4] https://www.tensorflow.org/

表 6.4: 年齢幅を制限した場合と制限しない場合の正解率の比較

範囲		URI freq	Word freq	Word cluster freq	Image hist freq
20 以上		0.3334	0.3408	0.3396	0.3397
20 から 59		0.3787	0.3805	0.3795	0.3796

6.7.4 Web コンテンツの抽出

Word Frequency Vectors と Word Cluster Frequency Vectors を構成するために，History Log に存在する 123,601 の URI に対して HTML のリクエストを送信した．その結果，52,874 件の Web ソースが得られた．残りに関しては，HTML のリクエストが拒否されたか，既に Web ページが閉鎖されていたために取得できなかったため，今回は HTML が取得できなかった URI が含まれたレコードは使用しないこととする．

上記のリクエスト後，Web ソースからテキストを抽出する．HTML タグをすべて取り除くために，HTML および XML を解析する Python ライブラリである *BeautifulSoup*[5] を使用している．また，HTML タグを取り除いた後，日本語の形態素解析を行うために形態素解析ライブラリである *MeCab*[6] を使用している．

上記とは別に，Image Histogram Vectors の構成のためにスクリーンショットを撮影する．Python から操作が可能な仮想ブラウザである *Selenium*[7] を使用して撮影を行った．History Log に存在する 2,754 のドメインに対して HTTP リクエストを送信し，5 秒待機したのち，そのページの最上部を解像度 800x600 で撮影した．

6.7.5 有効なデータのクリッピング

今回提供して頂いているデータの中にはエラーデータが含まれており，全データを使用すると正解率の低下を招くおそれがある．そのため，欠損値を含むレコードに関しては使用せずにモデルの構築を行う．また前述の通り，HTML のリクエストが拒否されるなどして使用できなかった URI を含むレコードもすべて使用しないこととする．

6.7.6 History Log のクレンジング

協力企業のアドバイスを受けて，実際の使用条件に合わせ，モデル構築に使用する年齢幅を制限した．6.4 で述べたとおり，今回のデータのメインボリュームは 40 歳前後である．また，この予測システムを協力企業が使用することを想定したとき，20 から 59 歳に関して予測が可能であれば実用に耐えうるというコメントを頂いたため，History Log のデータを 20 から 59 歳に絞ってモデル構築を行った．

[5] https://www.crummy.com/software/BeautifulSoup/
[6] http://taku910.github.io/mecab/
[7] http://www.seleniumhq.org/

これに関して，簡単な実験を行い，年齢幅に制限を設けた場合においても正解率が下がらないことを検証した．検証を簡単にするため，Random Forest のみで比較を行っている．表 6.4 にその結果を示す．この結果から，60 歳以上のデータを使用しないほうが正解率が上昇すると判断できるため，これ以降のモデル構築の操作においては 20 から 59 歳までのデータを用いる．

6.7.7 グリッドサーチによるパラメータ調整

Random Forest，XGBoost，ニューラルネットワークにはそれぞれモデル構築のためのパラメータが存在する．これらについて，グリッドサーチを行ってパラメータ探索を行った．なお，年齢については平均二乗誤差が最も少なくなるパラメータを，性別に関してはもっとも正解率が高くなるパラメータを採用することとした．

Random Forest におけるグリッドサーチ

以下の条件で Random Forest のパラメータのグリッドサーチを行っている．

- estimators: 構築する木の最大数．100，500 または 1000.
- min samples split: 木による分割の際，最低どれだけのデータ数を残すか．2 または 10.
- max depth: 木の深さの最大値．5 または 10.

XGBoost におけるグリッドサーチ

以下の条件で XGBoost のパラメータのグリッドサーチを行っている．

- n estimators: 構築する木の最大数．100，500 または 1000.
- max depth: 木の深さの最大値．5 または 10.

Grid-Search for Neural Network

以下の条件でニューラルネットワークのパラメータのグリッドサーチを行っている．今回の検証においてはすべての層において同じユニット数を使用することとしている．

- 層の数．2 層 (すなわち，入力層と出力層のみ) から 10 層まで 1 刻み.
- それぞれの層のユニットの数．32,64,128,256,512, または 1024.

表 6.5: グリッドサーチによって選ばれたパラメータ

	Random Forest		XGBoost		NeuralNetwork	
	estimators min samples split max depth		estimators max depth		layers units	
	Age	Sex	Age	Sex	Age	Sex
URI freq	1000 10 10	1000 2 10	500 5	500 5	2 256	5 32
Word freq	1000 10 10	1000 2 10	100 5	100 5	4 64	2 32
Word cluster freq	1000 10 10	1000 2 10	100 5	100 5	10 512	2 256
Image hist freq	1000 10 10	500 10 10	100 5	100 5	8 256	2 512

グリッドサーチの結果

表 6.5 に，グリッドサーチによるパラメータの選択結果を示す．なお，グリッドサーチにおける各パラメータの正解率と平均二乗誤差については付録に記す．Randam Forest のケースにおいては，ほぼ estimators が 1000，min samples が 10，max depth が 10 という結果となっている．ただし，Image Histogram Frequency の性別においてのみ，estimators が 500 となっている．

XGBoost のケースにおいては，URI Frequency を除くすべてで値が一致しており，estimators が 100，max depth が 5 という結果になっている．

ニューラルネットワークのケースにおいては，値に大きくばらつきが見られた．また，正解率が同じ値を示すパラメータも存在していたため (例えば，Word Cluster Frequency Vectors の性別予測における 2 層 256 ユニットと，6 層 512 ユニット．)，その場合は学習時間が短いほうのパラメータを採用した．

6.7.8 モデル構築と予測結果

4 つのベクトル化手法を用いて 3 種類のアルゴリズムを用いた場合の年齢と性別予測を行う．これまでの操作を経て 32,165 人のユーザのデータが有効であると判断された．このうち 24,124 件のデータを学習に，残りの 8,041 件をバリデーションに使用し，k-分割交差検証を行った．

表 6.6: 年齢予測の正解率

アルゴリズム		URI freq	Word freq	Word cluster freq	Image hist freq
Random Forest		0.3804	0.3829	0.3835	0.3815
XGBoost		0.3984	0.3898	0.3855	0.3890
Neural Network		0.3934	0.3961	0.3833	0.3795

表 6.7: 性別予測の正解率

アルゴリズム		URI freq	Word freq	Word cluster Freq	Image hist Freq
Random Forest		0.6380	0.6552	0.6914	0.6806
XGBoost		0.7015	0.7048	0.6940	0.6877
Neural Network		0.6982	0.7048	0.6902	0.6790

表 6.6 と表 6.7 に正解率を示す．年齢予測においては，XGBoost と URI Frequency Vectors を組み合わせた場合が最もよい正解率を示した．また，性別予測については，XGBoost と Word Frequency Vectors を組み合わせた場合，ニューラルネットワークと Word Frequency Vectors を組み合わせた場合の 2 パターンにおいて最良の結果となった．また，どちらの予測においても，XGBoost とニューラルネットワークの結果は拮抗している．

6.7.9 予測結果についての考察

ここでは 4 種類のベクトル化手法について結果を示した．ユーザの趣味嗜好が現れていると考えられる Word Frequency Vectors，Word Clusters Frequency Vectors が最も良い結果が現れると筆者は想定していたが，この結果を見ると，URI Frequency Vectors が良い結果を示すことがあることがわかった．なぜ URI を使用することが良いのか，今後よく考える必要がある．例えば，ユーザは URI をクリックする際にその先の Web ページにどのようなテキストや画像があるか判断していない，ということが考えられる．もしくは，年齢層ごとの偏りなどから，十分な学習データ量ではなかったかもしれなかったことが考えられる．

Image Histogram Frequency Vectors は他の手法と比べ若干低い正解率となってしまっている．今回のニューラルネットワークについては全結合層を用いているが，画像に対しては CNN が用いられることが通常であり，ネットワーク構造の再考の余地がある．

XGBoost とニューラルネットワークの結果については，前述の通り拮抗しており，どちらも高い精度が出ている．しかし，実用に際しては，精度のみを重視せずに Random Forest や XGBoost を採用したほうが良い場合もある．Random Forest や XGBoost などの手法は，ツリーを構築する際のデータを分割した理由を示すことができるからである．ニューラルネットワークを使用した場合，精度が良い場合でも，どんなメタデータが年齢および性別の予測に寄与したか判断が難しいこともある．したがって，実システムにこのモデルを搭載する場合は，実使用を想定して手法を選択する必要がある．

6.8 本章の結論

本章では，行動データとデモグラフィックデータを相互に接続するための新たなモデルである Action-Demographic Interconnection Model を提案し，その重要性について述べた．このモデルは外部データを接続することにより，行動データにメタデータを付与することができる．これは行動データとデモグラフィックデータの単純な関係を示すのみではなく，さらに深掘りして Web コンテンツとデモグラフィックデータの関係を示すものである．このモデルの実現によって，消費者の行動に合わせた Web コンテンツの生成が可能となる．

このモデルを実現するために，行動データをベクトル化する 4 つの手法を提案した．また，実際に Web アクセスログを使用し，4 つのベクトル化手法によるモデルを構築し，デモグラフィックデータ予測の検証を行った．

将来的には Web 上の行動履歴以外の行動データにモデルを適用して検証していくことを想定している．今回はテキストとシンプルな画像のみを検証したが，楽曲や映像に適用したり，それらが組み合わさったものにも適用可能であると考えられる．また，今回の予測においてアクセスログの時系列情報は使用されていないが，これを用いた予測を行うことにより，より精度の良い予測行うことが可能である考えられる．そのため，時系列データへ適用可能な拡張モデルを提案することが今後の課題となる．

第7章　オウンドメディアにおけるカスタマージャーニー分析

7.1　研究の背景

この章では，オウンドメディアのコンテンツに関するカスタマージャーニーの分析および，カスタマージャーニーに関する新たな評価指標の提案を行う．

現在，多くの企業が自社製品やサービスに関する Web サイトを運営し，様々なコンテンツを提供している．このような Web サイトをオウンドメディアと呼ぶ．企業がオウンドメディアを運営する役割は，新たなサービスの周知などにとどまらず，既存顧客の維持や，新規顧客の獲得といった側面を持っている．これらを効率よく行うためには，オウンドメディアのコンテンツに関する最適化が必要だと考えられる．最適化に関して，Web サイトの評価指標として代表的なものはページビューであるが，オウンドメディアにおいてページビューのみが重要であるとは限らない．ここではオウンドメディア内の様々なコンテンツに触れてもらうという目的を掲げ，ユーザのオウンドメディア内での滞留時間を評価基準とする．その滞留時間を増加させているコンテンツが優秀だとみなし，それを基準にコンテンツの最適化を行う．ここで述べるコンテンツとは，Web ページ上に掲載されたテキストやトピック，画像や楽曲，映像データを指す．

上記に述べるように，今回提案する手法は Web ページ単位での評価ではなく，Web ページ内のコンテンツを単位をした評価を行う．これにより，消費者にアプローチできるコンテンツを自動生成することが可能となる．

このような背景において，この章では Web ページのコンテンツの滞留時間に関する評価指標として User Trajectory Rank と User Retention Rank の 2 つを定義し，コンテンツを評価する．また，実際のオウンドメディアにおけるアクセスログを使用して，提案する評価指標の有効性を示す．

7.2　オウンドメディアの関連研究

7.2.1　メディアの種類

Corcoran[65] は Web マーケティングに関するメディアを 3 つの種類に分類した．それぞれの役割に応じてオウンドメディア，ペイドメディア，アーンドメディアと呼ばれている．Corcoran によるそれぞれのメディアの定義を表 7.1 に示す．

表 7.1: Corcoran による 3 つのメディアタイプ

Media type	Definition	Examples
Owned Media	Channel a brand controls	Web site, Mobile site, Blog, Twitter account
Paid Media	Brand pays to leverage a channel	Display ads, Paid search, Sponsorships
Earned Media	When customers become the channel	WOM, Buzz, “Viral”

オウンドメディア

マーケター (企業や広告代理店など) により管理されているメディアである [66]．オウンドメディアは，潜在的な顧客と企業の長期的な関係を構築する．オウンドメディアの例としては，企業が運営する Web サイトやブログ，SNS のアカウントがある．オウンドメディアの利点としては，管理がしやすく，低コストであり，長期的に運用が可能であることである．一方で，必ずしも良い効果を生むとは限らず，その効果が現れて観測するまでの期間も長い．

ペイドメディア

マーケターによって購入されたメディアである [66]．ペイドメディアはオウンドメディアの宣伝を行ったり，アーンドメディアを生成したりするための，2 つのメディアを取り持つ触媒のような役割をもつ．Web 広告の掲載や，テレビ番組やアスリートのスポンサーなどがこのメディアにあたる．ペイドメディアの利点は管理がしやすく，情報の伝播が早いことである．一方で名前が指し示すとおりコストがかかること，様々な方面にアプローチをかけなければならないこと，他のメディアに比べ，メディアの閲覧者の反応が弱いところが短所である．

Earned Media

アーンドメディアはマーケターが管理できず，購入することもできないメディアである [66]．アーンドメディアは時として，オウンドメディアやペイドメディアの効果を促進するはたらきがある．このメディアの例としては，口コミ，“バズる”，まとめサイトの記事などがある．これらは消費者にとってはもっとも信用できて透明性がある情報のように感じられ，企業側としても売上に大きく貢献するメディアである．しかし，再度述べるようにこれらのメディアは企業側が意図的にコントロールすることができないため，炎上といった形を取って悪い結果をもたらすことがある．また，アーンドメディアがもたらした効果がどれほどであったか測定することが難しいとされている．

7.2.2 カスタマージャーニー

カスタマージャーニーとは，その企業にとって顧客でなかった消費者が，何度もその企業のサービスを消費する消費者となるまでの過程を指す [67, 68]．また，ある消費者がある商品やサービスに関心を示した場合に絞って，その消費者が最終的な購入に関する意思決定に至るまでの様々な経験を指すこともある [69]．

表 7.2: アクセスログの形式

Tracking data	View date(yyyy-mm-dd hh-MM-ss)	URI
s01234567.hostA.ne.jp	2015-06-10 00:00:01	http://website.co.jp/page1
s02481357.hostB.ne.jp	2015-06-10 00:00:01	http://website.co.jp/page5
s01234567.hostA.ne.jp	2015-06-10 00:00:06	http://website.co.jp/page12
hogehoge.hostC.ne.jp	2015-06-10 00:00:10	http://website.co.jp/page4
⋮	⋮	⋮

7.3 データの統計的分析結果

この節では，あるオウンドメディアに関しての統計的分析結果を述べる．

7.3.1 分析対象となるオウンドメディア

今回は国内の日用品メーカーに協力していただき[1]，オウンドメディアの分析を行った．今回，協力企業が運営しているオウンドメディアは自社製品や日常生活に関する記事を 434 ページ公開している[2]．ただし，この数には他のページの目次のみが掲載されているページは含まない．これらの Web ページは “そのほかのおすすめ記事” として似たような内容を持つページへのリンクを張っている．

また，今回の分析にあたり，このオウンドメディア内のアクセスログを提供して頂いた．その例を表 7.2 に示す．それぞれのレコードはトラッキングデータ (ホストアドレス)，アクセス日時，URI の 3 つで構成されている．

トラッキングデータはユーザを一意に識別するためのホストアドレスである．アクセス日時はそのユーザがそのページにアクセスした日時である．URI はオウンドメディア内の閲覧したページのアドレスである．

7.3.2 オウンドメディア内でのユーザ遷移

今回のアクセスログは 393,906 人のユニークなユーザを含むログである．同一ホストアドレスによる短時間でのオウンドメディア内アクセスをユーザの遷移とし，各ユーザがオウンドメディア内のページをどれだけ閲覧してから他のメディアに移っているかを調査した．この短時間という時間については，今回の検証では 1 時間以内としている．

その結果が表 7.3 となる．実に 99.0% となる 39 万人弱のユーザは，検索エンジンなどからオウンドメディア内に流入してから，オウンドメディア内の他のページを閲覧することなく閲覧をやめてしまうことが判明した．

[1] NDA により企業名は非公開
[2] 分析当時．2015 年 8 月頃．

表 7.3: オウンドメディア内でのページ間移動回数

Times	Users		Times	Users		Times	Users		Times	Users
0	389792		10	9		21	1		40	1
1	2890		11	9		22	2		42	2
2	626		12	12		23	2		43	1
3	229		13	7		24	1		50	1
4	100		14	7		26	2		52	1
5	67		15	4		28	2		78	1
6	47		16	4		29	1		79	1
7	31		17	3		33	1		80	1
8	25		18	2		34	1			
9	17		20	1		35	2			
							Total users		393906	

7.3.3 オウンドメディア内での流入数と流出数

オウンドメディア内でお互いにリンクを張っていることがどれだけ効果を持っているのか，オウンドメディア内での各ページの流入数および流出数を調査した．ここでの流入数とは，ある特定のページに関して，それ以外の何種類のページから移動してきたかを指し，流出数はその反対に，ある特定のページから，どれだけのオウンドメディア内ページに移動させているかを示す．

図 7.1 は，流入数のグラフである．例えば，一番左の点はオウンドメディア内 110 種類のページから流入があるページである．これを見ると，10 種類以上のページから流入していないページが半分以上存在していることがわかる．

図 7.2 は，流出数のグラフである．こちらも，半分以上のページが，10 種類以上の他のページに遷移させていないことがわかる．

7.3.4 ユーザの遷移ネットワーク

図 7.3 は今回のオウンドメディア内でのユーザの遷移を表すネットワークである．5 人以上の移動がある場合にページ間を線でつないでいる．

最も大きいネットワークはスターネットワークを構成している．このスターネットワークの中央部に存在しているページは，このオウンドメディアにおいて多く閲覧されているページであることがわかる．

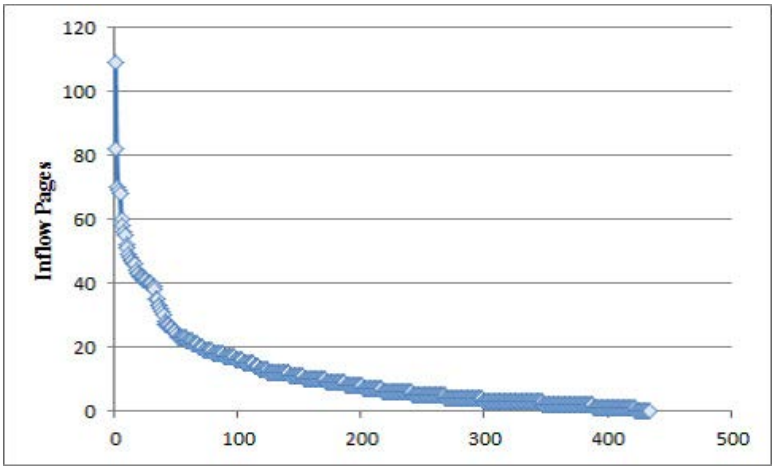


図 7.1: オウンドメディア内での流入ページ数

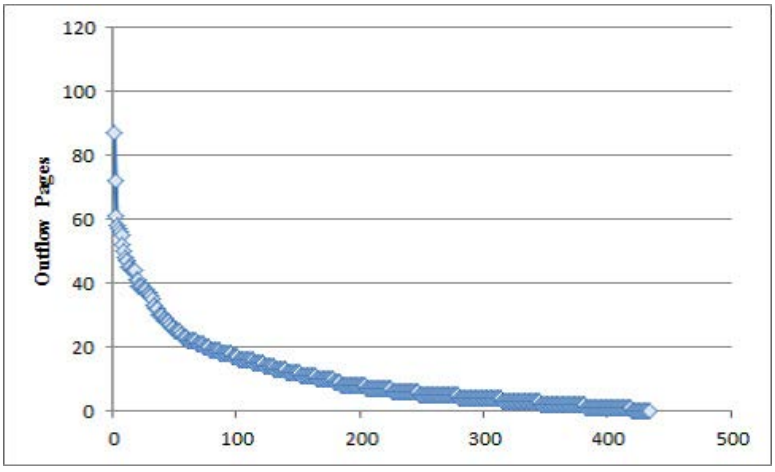


図 7.2: オウンドメディア内での流出ページ数

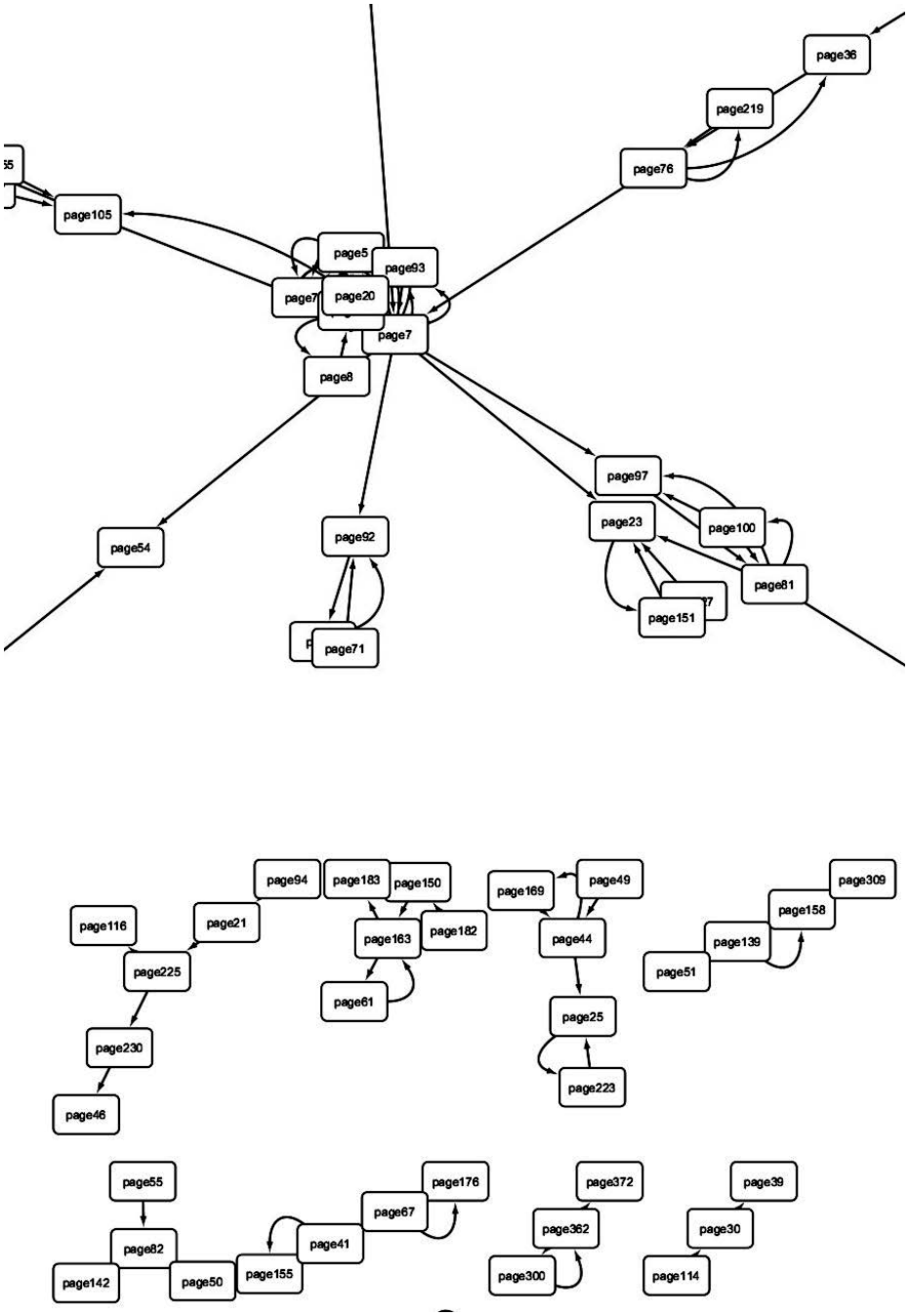


図 7.3: User movements on the Web-site of a certain Owned Media

7.4 オウンドメディアにおける評価指標の提案

ここでは，オウンドメディア内のカスタマージャーニーに関する評価指標を 2 つする．ユーザーが接触しているコンテンツの遷移に着目したコンテンツの重要度である User Trajectory Rank(UTR) と，ユーザーがオウンドメディアに滞留するかに着目したコンテンツの重要度である User Retention Rank(URR) を提案する．

UTR は，重み付き PageRank[70, 71] をベースとしている．ただし，PageRank は Web ページのリンク関係に着目しているが，UTR はユーザーの実際の遷移情報に着目して計算を行う．また，PageRank は Web ページ間の関係に着目しているが，UTR はさらに深掘りを行って，Web ページ内のコンテンツ間の関係に着目して計算を行う．ただし，今回は分析対象となるコンテンツに相当するものがオウンドメディア内の Web ページである．

7.4.1 User Trajectory Rank の計算

User Trajectory Rank は以下の手順で計算する．この評価指標は，ユーザーが接触しているコンテンツの遷移に着目したコンテンツの重要度を示す．

Step 1:コンテンツ間ユーザ遷移行列の生成

オウンドメディア内に n 種類のコンテンツが存在するとき，そのコンテンツ集合を $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ とする．このとき，コンテンツ間ユーザ遷移行列 A は n 次元正方行列となり，各列各行がコンテンツ集合 $\mathbf{c}$ の要素に対応する. 行列の各要素について，ユーザがコンテンツ c_i に接触してから c_j に接触した場合，$A_{i,j}$ にその接触を行ったユーザ数をカウントして格納する．そうでなければ $A_{i,j} = 0$ とする．

Step 2:重み付き PageRank の計算

コンテンツ間ユーザ遷移行列 A を用いて，重み付き PageRank の計算を行う．これによって得られた数値をそのコンテンツの User Trajectory Rank とする．

7.4.2 User Retention Rank の計算

User Retention Rank は以下の手順で計算する．この評価指標は，ユーザーがオウンドメディアに滞留するかに着目したコンテンツの重要度を示す．オウンドメディア外のページを考慮する open-UTR と，考慮しない closed-UTR の概念を取り入れる．

Step 1:closed-UTR の計算

オウンドメディア内に n 種類のコンテンツが存在するとき，そのコンテンツ集合を $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ とする．このコンテンツ集合に関して UTR を計算し，closed-UTR とする．closed-UTR は，オウンドメディア内におけるコンテンツの重要度を示す．

表 7.4: URR によるページの抽出結果 (Top 10)

URI	閲覧数	流入数	流出数	流出率	closed-UTR	open-UTR	URR
Page7	28604	109	87	0.9862	0.0374	0.0015	0.0413
Page20	16656	82	58	0.9902	0.0221	0.0014	0.0624
Page5	7591	49	38	0.9779	0.0148	0.0014	0.0963
Page53	7108	58	39	0.9852	0.0127	0.0014	0.1066
Page70	4685	39	28	0.9819	0.0108	0.0013	0.1180
Page17	8421	69	50	0.9891	0.0119	0.0015	0.1222
Page44	13403	60	61	0.9921	0.0117	0.0015	0.1290
Page105	11399	68	72	0.9881	0.0107	0.0014	0.1334
Page362	3306	70	52	0.9719	0.0117	0.0016	0.1356
Page163	2913	48	38	0.9715	0.0097	0.0013	0.1384
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Step 2:open-UTR の計算

オウンドメディア外のあらゆるコンテンツを 1 つのコンテンツとして捉え，c_o とする．そのオウンドメディア外コンテンツを加えたコンテンツ集合を **c'** $= (c_1, c_2, \ldots, c_n, c_o)$ とする．このとき，オウンドメディア外を考慮したコンテンツ間ユーザ遷移行列 A' は $n+1$ 次元正方行列となる．$n+1$ 行および $n+1$ 列以外の要素に関しては，通常の UTR と同様に数値を格納する．$n+1$ 列目に関しては，コンテンツ c_i に接触したあとオウンドメディア外 c_o に流出してしまっている場合，$A'_{i,n+1}$ にその流出したユーザ数を格納する．そうでなければ $A'_{i,n+1} = 0$ とする．この新たなコンテンツ間ユーザ遷移行列 A' に関して UTR を計算し，open-UTR とする．open-UTR が高いコンテンツは，外部にユーザを流出させる可能性があるコンテンツとみなせる．

Step 3:User Retention Rank の計算

以下の計算式によって User Retention Rank(URR) を計算する．

$$User\ Retention\ Rank = \frac{open\text{-}UTR}{closed\text{-}UTR}.$$

URR が小さいほど，ユーザの滞留に関してよいコンテンツであることを示す．

7.4.3 提供データに対する URR の適用

表 7.4 に URR の計算結果を示す. この評価指標においては *Page7* がもっとも良いコンテンツ (ページ) であると判断されている．特に *Page7*，*Page20*，*Page5* の 3 ページについては図 7.3 の中央部に配置されているページであり，感覚と一致する．

7.5 本章の結論

この章ではオウンドメディアのカスタマージャーニー分析に関する新たな評価指標である User Trajectory Rank と User Retention Rank を提案した．また，これらを実際のオウンドメディアに適用し，優秀なコンテンツを抽出できることを示した．

今回はページ間の関係性に着目して UTR と URR を計算したが，さらに Web ページ内のコンテンツを深掘りし，単語間の関連性や画像間の関連性，さらには異種メディア間の関連性を含めた URR の計算を行うことにより，Web ページのデザインまでを含めた自動コンテンツ生成を行うことが可能となる．また，異なるオウンドメディアについても適用し，有効性を検証する必要がある．

第8章　結論

本稿では人間とコンピュータが協調して要望を解決する枠組みとして，評価指標を複数搭載し，解決したい要望に応じて評価指標を選択するためのインターフェースである評価指標マネジメントインターフェースを提案するとともに，その具体的実現例を複数述べ，それらの検証を行った．

第 3 章では画像に関する評価指標マネジメントインターフェースである画像特徴量手法自動選択メタシステムを提案した．画像の研究者によって提案されてきた専門知である画像特徴量を Bag-of-Keypoints の投票の概念によって統合する仕組みを導入し，単一の画像特徴量を用いるよりも精度の良い分類を行うことが可能となった．また，入力された分類済み画像セットを入力すると，分類に関する評価指標が最適化されるような，統合された画像特徴量手法を備えた画像分類システムそのものを構築するメタシステムを提案した．

第 4 章では現実の事象をトリガーとし，これまでインターネット上に構築されてきたレガシーなメディアコンテンツをレガシーなシステムを通してリアルタイムに取り扱うことができる枠組みのメディアドリブンリアルタイムコンテンツマネージメントフレームワークを提案した．また，このフレームワークの具体的実現例として小中学校向け授業内発想支援システム *AI-Josyu* を協力企業と製作した．この授業内発想支援システムを実現するために，授業内の発言から自動的に重要な単語を選択するための評価指標である授業内単語重要度を提案した．

第 5 章ではテレビ CM の字コンテを入力とし，その CM を製作した際に視聴者が受ける印象を予測して出力する CM 字コンテ好感度予測システムの提案を行った．CM の好感度を数値で予測する学習済みニューラルネットワークと，単語と印象の相関を羅列した相関行列を備えた $\mathcal{ML}$ を構築することにより，CM の製作の前段階の時点で CM の好感度を予測することが可能となった．この $\mathcal{ML}$ は既存の $\mathcal{ML}$ と異なり，専門家の知識を用いて構築するのではなく，データから導かれた関係性を使用して作られた $\mathcal{ML}$ である．また，CM 関連企業と協力して上記システムを実装・構築し，*CREATIVE BRAIN* としてリリースを行った．

第 6 章では Web アクセスログに含まれる Web コンテンツを意味的に分析し，閲覧しているコンテンツと閲覧しているユーザの属性の関係性を明らかにする Action-Demographic Interconnection Model を提案した．また，提案モデルの実現に際し，Web コンテンツをベクトルして表現するための手法を 4 つ提案し，それぞれを採用したモデルにおけるデモグラフィックデータの予測の有効性を検証した．

第 7 章ではオウンドメディアのコンテンツを評価するための評価指標として，User Trajectory Rank と User Retention Rank の 2 つを提案した．User Trajectory Rank(UTR) はユーザーが接触

しているコンテンツの遷移に着目したコンテンツの重要度を示し，User Retention Rank(URR)はユーザーがオウンドメディアに滞留するかに着目したコンテンツの重要度を示す．また，実際のオウンドメディアに提案した評価指標を適用し，それぞれの評価指標の目的に応じたコンテンツを抽出した．

本稿では様々な具体的実現例を述べてきたが，これら具体的実現例が指し示すことは，データインテンシブサイエンスの潮流によって，研究のあり方が変わってしまったというところにある．これまで行われてきたような還元主義的な考え方からすれば，研究者，開発者，企画者，販売者などがそれぞれ最善を尽くして技術や製品を完成させ，それを組み合わせることによってビジネスというものは成り立つと考えられてきた．しかし，従来の研究手法に倣って研究者が最良だと思うものを作り，それを企業に持ち込んだとしても，それが企業にとって役に立つ技術になるとは限らないことが第 5 章の CM 字コンテ好感度予測システムの実現で強く示唆されており，現在のビジネスがそうであることを強く指し示している．この CM の研究においては好感要因の予測に関してニューラルネットワークを使用した方法が一番良い結果を示していた．そのため，順位内一致率という評価指標においては当然ながらニューラルネットワークが採用されて然るべきであった．しかし，実際にリリースされた製品においてはピアソンの積率相関係数を用いた行列による手法が採用されている．これは，製品を販売する企業やその製品を使用するユーザにとって，この手法による予測結果に説明性がある，納得感があるという点が重要視されたからである．現状ではこれら説明性や納得感などを評価指標として数値化することは難しいかもしれないが，今回提案した評価指標マネジメントインターフェースによるサイクルを回し，実際にやりたかった要望と評価指標に相違がないか，あるいはそもそもの要望に足りないところはなかったのかを検証することによって，明確な評価指標を定義することはできなくとも，要望に合致するモデルを選択することができるようになったといえる．これは，企業における意思決定を支援したことに他ならない．評価指標マネジメントインターフェースは人間とコンピュータの協調を促し，そのような意思決定を支援するものになりうるインターフェースである．

また，世の中の潮流としては，メディアコンテンツからメタデータを抽出する演算に対する逆演算である「生成系」を構成することが求められるようになってきている．CM の事例で言えば，字コンテを入力として好感要因が出力され，どのような好感要因を得られるかはわかったが，その好感要因が自分の狙っていたものと違っていたときにどのように字コンテを修正すればわからない，というユーザの意見が多く聞かれた．これは，字コンテを好感要因に変換する作用素の逆演算によって，上げたい好感要因に対する字コンテを出力する生成系を作れば解決される．また，デモグラフィックデータ予測の事例では，年齢と性別がわかっているとき，それらをターゲットとするコンテンツを自動的に生成するという使い方が考えられる．これも，年齢と性別を予測する順方向の演算に対する逆演算を構成することにより解決される．今回のシステムにおいてニューラルネットワークによるモデルが採用されないケースが存在したのは，ニューラルネットワークによる演算は逆演算を構成できないという一点に集中する．しかしニューラルネットワーク以外の手法を採用する場合においても，生成系を構成するには多数の困難がつきまとう．一般にメディアコンテンツから特徴を抽出して

言葉のメタデータを抽出する操作は次元数を下げていく操作である．この逆演算を構成する場合には次元数を大きいほうに変換する作用素を構成しなければならず，不良設定問題となる．これまでの研究においてはメディアコンテンツに応じた制約 (楽曲で言えば，人間が聞いて自然な和音だと感じる和音に制約する [6]，など)，統計的な制約 (多数の楽曲データからメロディの音の上昇と下降の割合を算出して，その割合に近づくように楽曲を構成する[6]，など) によって不良設定問題を完全ではないにしろ克服してきた．CM の事例では，相関係数の逆行列を構成することにより，上げたい好感要因に合致した単語を推薦することは可能である．しかし，形態素解析によって生じた「文章から単語への変換」，Word2Vec によるベクトル化によって発生した「単語の順番の捨象」などによって，字コンテを逆演算で構成することは不可能である．このような生成系の構築こそが，評価指標マネジメントインターフェースを通して評価指標を再定義するサイクルを十分に活用するために必要な要素だと考えており，今後の研究における課題である．

今回提案した手法のいくつかが製品としてリリースできたのは，評価指標というものはモデルごとに決められた正解率や誤差の少なさではなく，最終的に使用するユーザや販売する企業によって決められているというインサイトに基づいて，協力した各企業が共通認識 (コンテキスト) を持って製品開発を進められたことによるものである．データインテンシブサイエンスの時代を迎えた今，蓄積されたデータを活用し，企業の垣根を超えた，技術開発から製品の販売，反応調査に至るまで一貫したビジネスへの転換期が多くの企業に訪れている．本稿が示す評価指標マネジメントインターフェースはそのような転換の一助となるインターフェースであり，また，コンピュータの役割を一変させるインターフェースである．このようなインターフェースが人間が自分の要望を見つめ直すきっかけとなり，意思決定を加速させるものになるように，今後も改善を続けたい．

謝辞

本稿を執筆するにあたり，筑波大学 システム情報系 北川 高嗣教授には幅広い視点からのご助言を頂き，研究内容に対する多くのご指導を頂きました．ここに感謝致します．

慶應義塾大学環境情報学部 清木 康教授には，マルチデータベースシステムの観点から多くのご助言を頂きました．ここに感謝致します．

筑波大学 計算科学研究センター 北川 博之教授，システム情報系 伊藤 誠教授，今倉 暁准教授には本研究に対する多くのご助言やご提案を頂きました．ここに感謝致します．

武蔵野大学 データサイエンス学部学科長 中西 崇文准教授には研究へのアドバイスをはじめ，本稿を完成させるために多大なるご尽力をいただきました．ここに感謝致します．

筑波大学 システム情報系 櫻井 鉄也教授には画像特徴量手法自動選択メタシステムへのご助言をいただくとともに，所属研究室の先生として非常にお世話になりました．ここに感謝致します．

国際大学グローバル・コミュニケーション・センター リサーチアソシエイトの岡田 龍太郎氏には，研究内容への多くのご助言を頂きました．また，本稿に対し様々なご指導を頂きここに感謝致します．

株式会社サカワの佐川 寿忠氏をはじめとする皆さまには，*AI-Josyu* の構築の際に多くの議論を重ね，システムとしてのリリースを行うことができました．ここに感謝致します．

株式会社コラージュ・ゼロの小島 拓也氏をはじめとする皆さまには，*CREATIVE BRAIN* がより良いシステムとなるよう多くの議論をさせていただきました．ここに感謝致します．

これまでの学生生活において関わりのあった情報数理研究室の皆さま，国際大学グローバル・コミュニケーション・センターの先生方・スタッフ・リサーチアソシエイト・リサーチアシスタントの皆さま，日々支えて頂いたこと，ここに感謝致します．

付 録A　予備実験

A.1　クラスタ数と Visual-words 作成時間の関係の調査

画像特徴量手法自動選択メタシステムにおける Visual-words の作成時間を，クラスタ数依存で調査する．提案方式の検証を繰り返し行うことを考え，5 分程度で Visual-words の作成が完了するようなクラスタ数を決定する．

実験に使用したラップトップコンピュータのスペックを以下に記載する．

OS Windows 7 Professional 64-bit (6.1, Build 7601) Service Pack 1

Processor Intel(R) Core(TM) i5-3210M CPU @ 2.50GHz (4 CPUs), 2.5GHz

Memory 4096MB RAM

表 A.1，図 A.1 に実験結果を示す．このグラフから，おおよそクラスタ数と作成時間は比例していることが推測できる．この結果から，SIFT と SURF を合わせて 4 分程度で Visual-words の作成が終了する 1000 を選択した．

A.2　デモグラフィックデータ予測におけるグリッドサーチの詳細

グリッドサーチにおける各パラメータの正解率および平均二乗誤差 (Mean Squared Error, MSE) を示す．最も良い結果を表内に太字で示す．

表 A.1: クラスタ数と Visual-words 作成の時間の関係

クラスタ数	SIFT[ms]	SURF[ms]
1	183	89
2	570	390
5	1008	593
10	1695	1040
20	2891	1935
50	6804	4448
100	13341	8463
200	27316	17462
300	41168	26941
500	67718	46092
1000	132474	87971
2000	265785	174030
5000	698952	451536

表 A.2: URI Frequency における年齢の Random Forest パラメータ別 MSE

estimators	min samples split	max depth	
		5	10
100	2	86.551	84.391
100	10	86.556	84.384
500	2	86.558	84.368
500	10	86.571	84.361
1000	2	86.560	84.364
1000	10	86.562	**84.359**

表 A.3: Word Frequency における年齢の Random Forest パラメータ別 MSE

estimators	min samples split	max depth	
		5	10
100	2	84.963	82.092
100	10	84.958	82.087
500	2	84.966	82.082
500	10	84.963	82.063
1000	2	84.954	82.048
1000	10	84.952	**82.034**

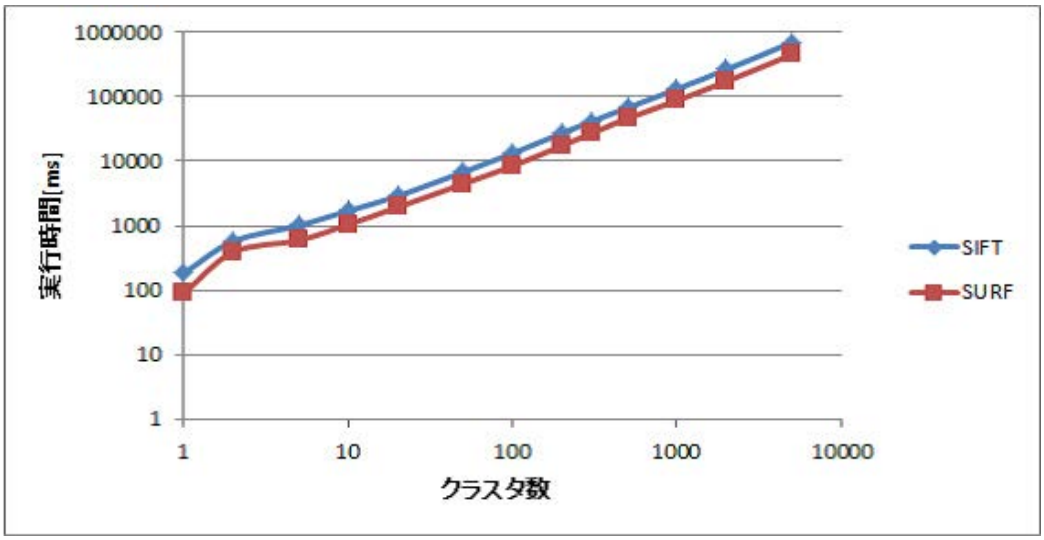


図 A.1: クラスタ数と Visual-words 作成の実行時間のグラフ

表 A.4: Word Cluster Frequency における年齢の Random Forest パラメータ別 MSE

estimators	min samples split	max depth	
		5	10
100	2	85.029	82.376
100	10	85.027	82.355
500	2	85.054	82.259
500	10	85.053	82.251
1000	2	85.036	82.250
1000	10	85.035	**82.232**

表 A.5: Image Histogram Frequency における年齢の Random Forest パラメータ別 MSE

estimators	min samples split	max depth	
		5	10
100	2	85.565	83.412
100	10	85.566	83.335
500	2	85.486	83.253
500	10	85.486	83.195
1000	2	85.474	83.239
1000	10	85.474	**83.188**

表 A.6: URI Frequency における性別の Random Forest パラメータ別正解率

estimators	min samples split	max depth	
		5	10
100	2	0.5731	0.6161
100	10	0.5732	0.6140
500	2	0.5710	0.6181
500	10	0.5714	0.6174
1000	2	0.5716	**0.6192**
1000	10	0.5718	0.6188

表 A.7: Word Frequency における性別の Random Forest パラメータ別正解率

estimators	min samples split	max depth	
		5	10
100	2	0.5994	0.6249
100	10	0.5995	0.6249
500	2	0.5961	0.6266
500	10	0.5955	0.6263
1000	2	0.5965	**0.6285**
1000	10	0.5718	0.6280

表 A.8: Word Cluster Frequency における性別の Random Forest パラメータ別正解率

estimators	min samples split	max depth	
		5	10
100	2	0.6461	0.6666
100	10	0.6457	0.6678
500	2	0.6509	0.6691
500	10	0.6510	0.6684
1000	2	0.6420	**0.6696**
1000	10	0.6419	0.6683

表 A.9: Image Histogram Frequency における性別の Random Forest パラメータ別正解率

estimators	min samples split	max depth	
		5	10
100	2	0.6387	0.6635
100	10	0.6382	0.6631
500	2	0.6391	0.6636
500	10	0.6389	**0.6646**
1000	2	0.6400	0.6642
1000	10	0.6405	0.6639

表 A.10: URI Frequency における年齢の XGBoost パラメータ別 MSE

estimators	max depth	
	5	10
100	80.910	80.436
500	**80.087**	82.079
1000	81.146	84.495

表 A.11: Word Frequency における年齢の XGBoost パラメータ別 MSE

estimators	max depth	
	5	10
100	**80.025**	80.433
500	80.749	83.993
1000	82.480	86.251

表 A.12: Word Cluster Frequency における年齢の XGBoost パラメータ別 MSE

estimators	max depth	
	5	10
100	**81.611**	83.202
500	83.103	86.394
1000	85.174	87.582

表 A.13: Image Histogram Frequency における年齢の XGBoost パラメータ別 MSE

estimators	max depth	
	5	10
100	**82.118**	85.693
500	83.253	85.889
1000	83.764	85.889

表 A.14: URI Frequency における性別の XGBoost パラメータ別正解率

estimators	max depth	
	5	10
100	0.6590	0.6728
500	**0.6838**	0.6809
1000	**0.6838**	0.6767

表 A.15: Word Frequency における性別の XGBoost パラメータ別正解率

estimators	max depth	
	5	10
100	**0.6841**	0.6772
500	0.6819	0.6770
1000	0.6811	0.6731

表 A.16: Word Cluster Frequency における性別の XGBoost パラメータ別正解率

estimators	max depth	
	5	10
100	**0.6747**	0.6713
500	0.6707	0.6661
1000	0.6683	0.6657

表 A.17: Image Histogram Frequency における性別の XGBoost パラメータ別正解率

estimators	max depth	
	5	10
100	**0.6711**	0.6661
500	0.6688	0.6645
1000	0.6688	0.6657

表 A.18: URI Frequency における年齢の NN パラメータ別 MSE

units	layers								
	2	3	4	5	6	7	8	9	10
32	89.533	82.008	80.229	84.895	87.492	91.824	88.810	96.305	92.990
64	81.924	80.196	79.690	81.005	82.393	88.398	92.041	87.541	87.126
128	79.988	80.388	80.968	80.077	81.992	84.135	82.410	83.682	86.675
256	**79.262**	81.656	83.375	82.339	83.427	81.459	81.801	80.474	82.232
512	79.993	82.994	87.301	84.456	82.198	81.058	82.457	81.501	84.481
1024	81.459	88.045	91.306	85.348	83.453	84.444	82.151	82.520	81.746

表 A.19: Word Frequency における年齢の NN パラメータ別 MSE

units	layers								
	2	3	4	5	6	7	8	9	10
32	98.205	85.005	89.009	90.963	93.134	107.737	90.441	98.336	105.912
64	85.827	82.662	**81.695**	85.553	94.389	91.817	100.060	94.918	90.662
128	82.383	85.431	84.605	86.726	88.236	91.789	86.790	88.776	92.503
256	82.504	87.765	86.353	84.443	82.613	83.682	87.498	89.843	91.889
512	84.121	91.365	89.738	87.655	88.880	87.173	87.749	83.213	88.480
1024	86.306	93.058	94.267	93.788	90.857	91.784	86.642	89.331	88.165

表 A.20: Word Cluster Frequency における年齢の NN パラメータ別 MSE

units	layers								
	2	3	4	5	6	7	8	9	10
32	92.786	93.026	87.534	88.031	95.502	94.800	93.241	93.830	101.027
64	85.753	83.876	79.379	80.950	88.040	88.089	94.547	92.147	86.746
128	82.443	81.770	80.948	84.456	79.040	85.688	81.330	86.005	82.122
256	80.369	79.490	79.748	80.155	79.270	81.113	80.015	78.964	86.911
512	80.030	79.537	85.686	81.154	80.419	82.860	79.953	79.226	**78.554**
1024	79.454	80.413	82.218	79.306	80.195	82.070	91.048	81.085	79.756

表 A.21: Image Histogram Frequency における年齢の NN パラメータ別 MSE

units	layers								
	2	3	4	5	6	7	8	9	10
32	92.504	83.018	85.711	88.310	88.174	87.921	90.611	86.963	99.198
64	85.003	82.224	84.468	86.236	84.952	86.625	86.276	89.022	86.413
128	84.221	81.339	89.400	81.294	82.270	85.787	83.440	83.028	82.544
256	82.261	82.805	81.968	80.763	81.585	80.644	**80.409**	83.236	83.932
512	81.862	82.173	82.436	80.438	80.722	80.572	81.082	81.769	80.430
1024	81.271	81.550	82.822	81.632	80.437	80.700	88.315	80.569	81.151

表 A.22: URI Frequency における性別の NN パラメータ別正解率

units	layers								
	2	3	4	5	6	7	8	9	10
32	0.6971	0.6997	0.7022	**0.7039**	0.6781	0.6883	0.5547	0.6691	0.5405
64	0.7013	0.6978	0.6998	0.6944	0.6993	0.6750	0.6784	0.6021	0.6408
128	0.6995	0.6973	0.6963	0.6987	0.7028	0.6849	0.6884	0.6635	0.6701
256	0.6995	0.7027	0.7008	0.6876	0.6960	0.6982	0.6745	0.6869	0.6750
512	0.7006	0.7026	0.6983	0.6986	0.7016	0.7024	0.7028	0.6936	0.6980
1024	0.6888	0.7034	0.7013	0.6973	0.6962	0.6911	0.6773	0.6942	0.6990

表 A.23: Word Frequency における性別の NN パラメータ別正解率

units	layers								
	2	3	4	5	6	7	8	9	10
32	**0.7081**	0.7041	0.7069	0.7005	0.6912	0.6952	0.6999	0.6523	0.6727
64	0.7030	0.7012	0.7039	0.7067	0.6939	0.6753	0.6976	0.6419	0.6428
128	0.7052	0.7043	0.7028	0.7010	0.7003	0.6957	0.6808	0.6771	0.6530
256	0.7054	0.7041	0.6988	0.7019	0.7043	0.7007	0.6932	0.6846	0.6800
512	0.7030	0.6988	0.7019	0.6990	0.7001	0.7008	0.6974	0.6817	0.6901
1024	0.6990	0.6985	0.7049	0.6979	0.6981	0.6919	0.6948	0.6988	0.6961

表 A.24: Word Cluster Frequency における性別の NN パラメータ別正解率

units	layers								
	2	3	4	5	6	7	8	9	10
32	0.6928	0.6829	0.6955	0.6853	0.6928	0.6836	0.6097	0.6035	0.5493
64	0.6939	0.6931	0.6970	0.6970	0.6935	0.6815	0.5708	0.5949	0.6681
128	0.7001	0.6886	0.6981	0.6866	0.6920	0.6880	0.6893	0.6787	0.6557
256	**0.7010**	0.6917	0.6992	0.6984	0.6937	0.6882	0.6829	0.6858	0.6924
512	0.6970	0.6993	0.6899	0.6911	**0.7010**	0.6928	0.6961	0.6942	0.6972
1024	0.6924	0.6953	0.6857	0.6891	0.6884	0.6911	0.6889	0.6902	0.6866

表 A.25: Image Histogram Frequency における性別の NN パラメータ別正解率

units	layers								
	2	3	4	5	6	7	8	9	10
32	0.6844	0.6781	0.6848	0.6831	0.6801	0.5852	0.6532	0.6676	0.6682
64	0.6857	0.6867	0.6812	0.6858	0.6846	0.6759	0.6801	0.6784	0.5437
128	0.6868	0.6851	0.6819	0.6777	0.6791	0.6748	0.6871	0.6527	0.6846
256	0.6867	0.6751	0.6863	0.6748	0.6793	0.6636	0.6877	0.6626	0.6843
512	**0.6884**	0.6772	0.6873	0.6856	0.6817	0.6774	0.6748	0.6876	0.6707
1024	0.6702	0.6777	0.6858	0.6803	0.6745	0.6661	0.6610	0.6683	0.6767

参考文献

[1] Tony Hey, Stewart Tansley, and Kristin Tolle. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 2009.

[2] Takashi Kitagawa and Yasushi Kiyoki. Fundamental framework for media data retrieval systems using media-lexico transformation operator - in the case of musical midi data. *Information Modelling and Knowledge Bases XII*, pp. 316–326, 2001.

[3] 清木康. 感性や意味を計量するデータベースシステム-人間と情報システムの記憶系について-. *Keio SFC journal*, Vol. 13(2), pp. 19–26, 2013.

[4] Ryotaro Okada, Takafumi Nakanishi, and Takashi Kitagawa. A method of knowledge creation and knowledge utilization by generalized inverse operator. In *Proceedings of 2014 IIAI 3rd International Conference on Advanced Applied Informatics*, pp. 253–258, 2014.

[5] 吉野太智, 高木秀幸, 清木康, 北川高嗣. 楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用. 情報処理学会研究報告. DBS, データベースシステム研究会報告, Vol. 116, No. 2, pp. 109–116, jul 1998.

[6] 岡田龍太郎, 中西崇文, 本間秀典, 北川高嗣. メディアコンテンツを対象とした統計的一般化逆作用素構成方式とその楽曲メディアコンテンツ生成への適用. 情報処理学会論文誌, Vol. 57, No. 5, pp. 1341–1354, may 2016.

[7] Takashi Kitagawa, Takafumi Nakanishi, and Yasushi Kiyoki. An implementation method of automatic metadata extraction method for image data and its application to a semantic associative search. In *Information Processing Society of Japan Transactions on Databases*, Vol. 35(SIG12,TOD16), pp. 38–51, 2002.

[8] 柳井啓司. キーワードと画像特徴を利用した www からの画像収集システム. 情報処理学会論文誌. データベース, Vol. 42, No. 10, pp. 79–91, sep 2001.

[9] 柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌. コンピュータビジョンとイメージメディア, Vol. 48, No. 16, pp. 1–24, nov 2007.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the 2015*

IEEE International Conference on Computer Vision (ICCV), ICCV 15, pp. 1026–1034, USA, 2015. IEEE Computer Society.

[11] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, pp. 91–110, 2004.

[12] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, Vol. 110, No. 3, pp. 346–359, June 2008.

[13] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *International Conference on Computer Vision*, Barcelona, 11/2011 2011.

[14] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.

[15] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988.

[16] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ECCV'06, pp. 430–443, Berlin, Heidelberg, 2006. Springer-Verlag.

[17] A. Haar. Zur theorie der orthogonalen funktionensysteme. (erste mitteilung). *Mathematische Annalen*, Vol. 69, pp. 331–371, 1910.

[18] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, pp. 778–792, Berlin, Heidelberg, 2010. Springer-Verlag.

[19] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, Vol. 3, pp. 1157–1182, March 2003.

[20] 長畑香奈栄, 金沢靖. 適応的な特徴量選択による画像間の高精度な対応付け法の検討. *Information Processing Society of Japan (IPSJ) SIG Notes. Computer Vision and Image Media*, Vol. 24, , feb 2014. in Japanese.

[21] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA, 1999.

[22] Eric Nowak, Frédéric Jurie, and Bill Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part IV*, ECCV'06, pp. 490–503, Berlin, Heidelberg, 2006. Springer-Verlag.

[23] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pp. 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[24] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, Vol. 106, No. 1, pp. 59–70, April 2007.

[25] Ossama Abdel-Hamid, Abdel rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 10, pp. 1533–1545, 2014.

[26] Takashi Kitagawa, Takafumi Nakanishi, and Yasushi Kiyoki. An implementation method of automatic metadata extraction method for music data and its application to semantic associative search. *Systems and Computers in Japan*, Vol. 35, No. 6, pp. 59–78, 2004.

[27] Takafumi Nakanishi, Ryotaro Okada, and Takashi Kitagawa. Automatic media content creation system according to an impression by recognition-creation operators. In *Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6, 2016.

[28] Vivien Arief Wardhany, Sritrusta Sukaridhoto, and Amang Sudarsono. Indonesian automatic speech recognition for command speech controller multimedia player. *EMITTER International Journal of Engineering Technology*, Vol. 2, No. 2, pp. 39–48, 2014.

[29] Keith Bain, Sara Basson, Alexander Faisman, and Dimitri Kanevsky. Accessibility, transcription, and access everywhere. *IBM Systems Journal*, Vol. 44, No. 3, pp. 589–604, 2005.

[30] Rohit Ranchal, Teresa Taber-Doughty, Yiren Guo, Keith Bain, Heather Martin, J. Paul Robinson, and Bradley S. Duerstock. Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies*, Vol. 6, No. 4, pp. 299–311, 2013.

[31] Joseph Robison and Carl Jensema. Computer speech recognition as an assistive device for deaf and hard of hearing people, challenge of change: Beyond the horizon. In *Proceedings of Seventh Biennial Conference Postsecondary Education Persons Who Are Deaf or Hard of Hearing*, Vol. 6(4), pp. 299–311, 2013.

[32] Sandy Watson and Linda Johnston. Assistive technology in the inclusive science classroom: Devices and services can help science students with a wide variety of needs. *The Science Teacher*, Vol. 74, No. 3, pp. 34–38, 2007.

[33] Mike Wald. Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time. In *Proceedings of 10th International Conference on Computers Helping People with Special Needs ICCHP 2006*, pp. 683–690, 2006.

[34] Petr Červa, Jindrich Zdánský Jan Silovský, J. Nouza, and Jiří Málek. Real-time lecture transcription using asr for czech hearing impaired or deaf students. In *Proceedings of INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, Portland, OR, USA, 2012.

[35] Tatsuya Kawahara, Norihiro Katsumaru, Yuya Akita, and Shinsuke Mori. Classroom note-taking system for hearing impaired students using automatic speech recognition adapted to lectures. In *Proceedings of INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, pp. 626–629, Makuhari, Chiba, Japan, 2010.

[36] Margaret Martyn. Clickers in the classroom: An active learning approach. *EDUCAUSE Quarterly Magazine*, Vol. 30, No. 2, pp. 71–74, 2007.

[37] April R. Trees and Michele H. Jackson. The learning environment in clicker classrooms: student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology*, Vol. 32, No. 1, pp. 21–40, 2007.

[38] Motoki Yokoyama, Yasushi Kiyoki, and Tetsuya Mita. A similarity-ranking method on semantic computing for providing information-services in station-concierge system. *EMITTER International Journal of Engineering Technology*, Vol. 5, No. 1, pp. 16–35, 2017.

[39] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)*, pp. 746–751, 2013.

[40] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *Computing Research Repository (CoRR)*, 2013.

[41] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26*, pp. 3111–3119, 2013.

[42] CM INDEX 編集部. CM 好感度データブック 2018. 株式会社東京企画, 2018.

[43] Yasushi Kiyoki, Takashi Kitagawa, and Takanari Hayama. A metadatabase system for semantic image search by a mathematical model of meaning. *SIGMOD Record*, Vol. 23, No. 4, pp. 34–41, 1 1994.

[44] 本間秀典, 中西崇文, 北川高嗣. 任意の言葉を対象とした音韻印象変換作用素の構成とその感性検索への適用. 情報処理学会論文誌, Vol. 51, No. 5, pp. 1294–1309, may 2010.

[45] Takafumi Nakanishi and Keisuke Tamaru. An impression estimation and visualization method for tv commercials. In *2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pp. 1–6, Aug 2017.

[46] 佐藤真由美, 滝山桂子, 益本仁雄. テレビ放映された食品 cm の実態分析 : 機能的ベネフィットと心理的ベネフィットに着目して. 日本家政学会誌 = Journal of home economics of Japan, Vol. 54, No. 10, pp. 855–866, oct 2003.

[47] 野澤智行. タレント・キャラクターがテレビＣＭ認知および評価に及ぼす影響. 広告科学, Vol. 40, pp. 93–99, 2000.

[48] 浅川雅美, 岡野雅雄. テレビ cm に対する視聴者反応の分析–自由回答文のテキストマイニング. 文教大学女子短期大学部研究紀要, Vol. 48, pp. 1–6, jan 2005.

[49] 吉田知世, 小林一郎. 感情極性に基づく文書の俯瞰分析への取り組み. 第 73 回全国大会講演論文集, Vol. 2011, No. 1, pp. 387–388, mar 2011.

[50] Peter D. Turney and Michael L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, Vol. 21, pp. 315–346, 10 2003.

[51] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol. 12, No. 3, pp. 203–222, 2005.

[52] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 語釈文を利用した「p/n 辞書」の作成. 言語・音声理解と対話処理研究会, Vol. 33, pp. 45–50, 2001.

[53] George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, Vol. 38, No. 11, pp. 39–41, November 1995.

[54] Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. Using WordNet to measure semantic orientations of adjectives. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA).

[55] 小林重順. カラーイメージスケール 改訂版. 日本カラーデザイン研究所, 2001.

[56] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報処理学会研究報告. NL, 自然言語処理研究会報告, Vol. 161, pp. 89–96, may 2004.

[57] H. Cohen and C. Lefebvre. *Handbook of Categorization in Cognitive Science*. Elsevier, 2005.

[58] L.Breiman. Random forests. *Machine Learning*, Vol. 45(1), pp. 5–32, 2001.

[59] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[60] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Publishing Company, Incorporated, 2008.

[61] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Vol. 46, No. 3, pp. 175–185, 1992.

[62] Harry Zhang. The optimality of naive bayes. In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, (FLAIRS 2004)*, Vol. 2, 2004.

[63] Tsukasa Ishigaki, Takeshi Takenaka, and Yoichi Motomura. Category mining by heterogeneous data fusion using pdlsi model in a retail service. In *Proceedings of 2010 IEEE 10th International Conference on Data Mining (ICDM)*, 2010.

[64] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, Vol. 42, No. 1-2, pp. 177–196, 2001.

[65] Sean Corcoran. Defining earned, owned and paid media. http://blogs.forrester.com/interactive_marketing/2009/12/defining-earned-owned-and-paid-media.html, posted on Dec. 16, 2009, accessed on Jan. 19, 2016.

[66] Richard Hanna, Andrew Rohm, and Victoria L. Crittenden. We're all connected: The power of the social media ecosystem. *Business Horizons*, Vol. 54, No. 3, pp. 265–273, 2011.

[67] Martin Christopher, Adrian Payne, and David Ballantyne. *Relationship marketing : Bringing quality, customer service and marketing together*. Oxford, Butterworth-Heinemann, 1991.

[68] Ian Gordon. *Relationship Marketing: New Strategies, Techniques and Technologies to Win the Customers You Want and Keep Them Forever.* John Wiley & Sons Canada, 1998.

[69] Lawrence Ang and Francis A Buttle. Roi on crm: a customer-journey approach. In Peter J. Batt, editor, *Proceedings of the Inaugural Meeting of the IMP Group in Asia.*, pp. 1–20. IMP Group, 2002.

[70] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. *Technical Report of Stanford InfoLab*, 1999.

[71] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*, Brisbane, Australia, 4 1998.

論文目録

[査読有り論文誌]

(1). Kyohei Matsumoto, Takafumi Nakanishi, Toshitada Sakawa, Kengo Onodera, Shinichiro Orimo, and Hiroyuki Kobayashi, **AI-Josyu: Thinking Support System in Class by Real-time Speech Recognition and Keyword Extraction,** *EMITTER International Journal of Engineering Technology* Vol.7, No.1, pp. 366-383, 2019.

[査読有り国際会議論文]

(2). Kyohei MATSUMOTO, Ryotaro OKADA, Takafumi NAKANISHI, and Takashi KITAGAWA, **The method of image feature selection for integration of image classification by Bag-of-Keypoints,** *In Proceedings of 2015 International Conference on Computational Science and Computational Intelligence – CSCI-ISAI 2015, CSCI5084-ISAI,* pp. 589-594, Las Vegas, Nevada, USA, Dec. 2015.

(3). Kyohei MATSUMOTO, Takafumi NAKANISHI, and Takashi KITAGAWA, **Evaluation Indexes of Customer Journey for Contents of Owned Media,** *In Proceedings of the 26th International Conference on Information Modelling and Knowledge Bases - EJC 2016,* Tampere, Finland, pp. 395-402, Jun. 2016.

(4). Kyohei MATSUMOTO, Takafumi NAKANISHI, and Takashi KITAGAWA, **Semantic-Dependent Access Log Analysis for Predicting the Demographic Data,** *In Proceedings of the 28th International Conference on Information Modelling and Knowledge Bases - EJC 2018,* Riga, Latvia, pp. 110-129, Jun. 2018.

(5). Takafumi Nakanishi, Kyohei Matsumoto, Toshitada Sakawa, Kengo Onodera, Shinichiro Orimo, and Hiroyuki Kobayashi, **Media-driven Real-time Content Management Framework and its Application to In-Class Thinking Support System,** *In Proceedings of the 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC),* Surabaya, East Java, Indonesia, pp. 274-279, 2018.

(6). Takafumi Nakanishi, Kyohei Matsumoto, Toshitada Sakawa, Kengo Onodera, Shinichiro Orimo, and Hiroyuki Kobayashi, **A Class Content Summary Method based on Media-driven Real-time Content Management Framework,** *In Proceedings of 8th International Congress on Advanced Applied Informatics (IIAI-AAI 2019),* pp.795-798, 2019.