

第7班 共同研究

『インターネット・エコミュージアムのためのデータマイニングと ユーザインタフェース等の基盤技術に関する研究』

木下宏揚 佐野賢治 能登正人

森住哲也 宮田純子 小松大介

KINOSHITA Hirotsugu SANO Kenji NOTO Masato

MORIZUMI Tetsuya MIYATA Sumiko KOMATSU Daisuke

1 研究目的

本共同研究は、非文字資料を研究者間および専門家以外の人との間で情報の提供、共有などを行うために必要な基盤技術を構築し、実際の資料や研究者などを対象とした実証システムにより、その有効性を検証することを目的とする。第二期共同研究計画における非文字資料の検索、流通等に関する成果を踏まえて、第三期共同研究では、インターネット・エコミュージアムや只見町に開設予定の民俗博物館において必要なデータマイニングやデータの入力や検索に適したユーザインタフェースなどの基盤技術を開発することを目的とする。具体的な事業内容は以下の項目から成る。

1. オントロジーを用いたデータマイニングの実際の資料に対する応用
2. 資料のデータベースをクラウド化する際の個人情報保護と著作権管理
3. 資料の整理とデータ入力や流通を円滑化するためのゲーム理論や群知能などに基づく価値交換モデルの構築
4. 資料を直観的に取り扱うことを可能とする、群知能を用いた操作のコンテキストに基づくユーザインタフェースの構築

2 活動経過

2.1 2014年度研究経過

(1) 民具データベースのRDF化とオントロジーを導入した情報検索システム

近年研究や教育に役立てることを目的として、収蔵資料をデータベース化してウェブ上で公開する博物館が増加している。しかしこの博物館資料のデータベースが関係データベースに代表される従来の技術で実現される場合、資料の分類の変更や資料について熟知していない者の利用に十分に対応することはできない。また現状では博物館によって資料の整理の仕方が異なるため、複数の博物館のデータベースの情報を統合したり横断的に検索したりすることは極めて困難である。以上の問題を改善する試みとして、本研究では福島県只見町の民具データベースを対象にオントロジーを導入した情報検索システムを提案した。提案システムではデータベースはRDF化され、その上でデータの意味を

扱うのに必要となる語彙と知識が OWL によってオントロジーに記述されている。こうすることでデータベースに記載された資料に関するメタデータをコンピュータが理解できるようになるため、意味検索や他のデータベースとの相互運用の簡易化が可能になった。

(2) 個人情報保護のためのハイブリッドハニーポットと著作権管理のための電子透かし

情報技術の進歩は、深刻化が進んでいる様々なセキュリティ問題と連携してきている。様々なタイプのセキュリティシステムが絶えず進化してきている。例えば、侵入検知システム (IDS) はネットワーク内の攻撃を検知をし、ハニーポットは不法侵入を観察するのに効果的な方法である。我々はこれらのシステムを用いることで、ネットワーク内の統計的な情報と、ハッカーのハッキング技術を得ることができ、システムを強化することができる。本稿ではハニーポットと IDS とファイアウォールの技術の詳細を論じ、それらを動的に連携する手法を提案した。電子透かしでは、透かし情報を埋め込む際の手順を知る者にとっては、埋め込まれた情報を読み取り、改ざんすることができる。そのため、容易に透かし情報の読み取り、改ざんをできなくする必要がある。本研究では、二回ウェブレット変換を行い中間領域へスペクトル拡散を用いて埋め込むことでより高い秘匿性、および雑音に強い耐久性を備える電子透かしを提案した。

(3) 破産ゲームを用いた価値交換システム

コミュニティ内での価値の交換システムを提案した。研究室を例にし全体提携することによってより良い配分になるような結果を求めた。使用するゲーム理論によって配分の結果が異なり、モデルに合わせて使用することにより応用することができると考えられる。また先行研究と比べ、解が一つになるため極端な答えが出ることはない。今後の課題としてはプレイヤーの人数を増やして価値の交換を行う。また価値の交換範囲を研究室だけでなくもっと大きい範囲に拡大する。これにより能力の異なる研究者のグループに対して指導的立場の研究者がどのくらいのエフォートを割けば最もメンバーの満足度が高まるかが分かる。

(4) 相互類似による引力と斥力を表現した群知能を用いた情報リソースの管理

データの分野やジャンルなどの類似度に応じてファイルを集めることで、多くのファイルを一目で把握可能とするファイルシステムを提案した。この提案システムでは、ファイルは自己組織化された集合値と見なした群知能を用いることで、効率的にファイルを集めることを可能とする。さらに、古い情報をデジタル化した民具データを用いて提案システムによるシミュレーションをすることで、有効性を示した。

2.2 2015年度研究経過

(1) デジタルアーカイブ作成を前提とした民具データベースの構築

近年では資料情報のデジタル化が進みインターネット上で資料を検索し、参照可能な施設も増えつつある。しかし国内において、デジタル化された資料情報について保存方針は未整備である。そのため、情報を共有するにあたって互換性の確保は大きな課題である。加えて、博物館の資料は多様であり、デジタル化においては各博物館によって、あるいは資料群によって異なる規格で作成している。特に民俗学の分野では、研究者によって資料の分類方法は一様ではない。また、名称に方言を含むことが多く、類義語から記載のずれが生じやすいという問題がある。本研究では、民俗資料特有の情報

の維持と資料情報の互換性の確保を目的とし、民俗資料の一例として福島県南会津郡只見町に伝わる民具を対象とする。民具の持つ重要な要素に注目し、民具情報構造化モデルを提案した。また、民具のデータベース化を行うとともに、民具の検索システムについても検討を行った。

(2) 資料のデータベースの個人情報保護と著作権保護

近年、企業や個人が扱う情報は増加しており、その情報が漏えいする可能性も増加している。日本では2016年から「マイナンバー制度」が開始され、さらなる情報漏えいが懸念される。マイナンバー制度では情報が紐付けされているため、一つの情報から多くの情報が流出してしまう恐れがある。そこで、ハイパーグラフによる推論経路分析にロールベースアクセス制御モデルの「役割」という主体を制約条件として付加することで、分析の精度向上ができると考え、ハイパーグラフによる推論経路分析を、主体と客体の両面から評価するセキュリティモデルを提案した。著作権保護でデジタルコンテンツに電子透かしが用いられる。そして、デジタルコンテンツの劣化が少ない電子透かしの様々な研究が行われている。しかし、埋め込み対象として画像全体や全ての輪郭に埋め込んでいる方式が多い。よって、劣化を引き起こしている。本研究では、一部分に埋め込んで劣化を抑えたいと考え、離散コサイン変換とハフ変換を用いて、周波数領域に埋め込む方法を提案した。

(3) 資料の作成や流通のための価値交換システム

近年、ビットコインの原理が発表されて以来、ビットコインを利用する取引の増加も続いている。このような状況の中では、ビットコインの匿名性による犯罪が深刻化している。現在、取引を可視化するため、ビットコインの上位レイヤーのカラードコインに関する研究が進行中である。本研究は、カラードコインとアントコロニー最適手法(ACO)を用いて、購買者の商品購買行動を分析し商品に対する市場の好みを推測して商品の購入する順番を推薦するシステムを提案した。本システムは、カラードフェロモンという利用者の属性に相当するものを使い、属性が対応させた商品の購入順を無色フェロモンのACOによる最短経路問題に還元する。これにより、利用者は効率のいい買い物ができると考える。

(4) ユーザインタフェース

近年、個人で複数のデバイスを所持する人が増えてきており、それに伴い管理するファイルの量が増大している。そのため管理の手間が増し、扱いにくくなっている。ファイルの多次元の空間的な管理は、システム上要素が多いためPCは管理しやすいが、人間が把握するのは困難である。そこでPCの階層構造でのファイル管理を前提とし、それに対してファイルが使用される振る舞いを組み合わせる。本稿では、振る舞いをFA(Firefly Algorithm)を用いて評価し、ファイルの配置の最適化がなされる動的なファイルマネージャを提案した。

2.3 2016年度研究経過

(1) ブロックチェーンを応用した個人情報保護

ブロックチェーンはビットコインの主要な要素技術であり、信頼できる第三者を仮定しないで、順序関係に意味のある事象の系列を保証し、改ざんを防止する技術である。情報漏えい検出法における推論規則の発火条件をブロックチェーンによるログシステムを応用して管理する手法を提案した。これにより重要度の異なる研究資料や個人情報を入手する際の優先順序を示唆することが可能となる。

(2) 電子透かしを用いたペーパーウォレット

ビットコインアドレスとビットコインウォレットの秘密鍵を記録した物理的なペーパーウォレットを電子透かしを用いて構成する方法を提案した。これにより、研究者間のブロックチェーンベースの価値交換のセキュリティの向上が可能となる。

(3) ゲーム理論を用いた地域活性化システム

地域の活性化のためにプロジェクトを企画した場合、コストの負担に積極的なグループと消極的なグループが存在する。各グループ間の戦略を提携ゲームで、グループ内の提携を非定型ゲームでモデル化しプロジェクトが成功した場合と失敗した場合に対して、それぞれ最適なモデルとして浄化ゲームと破産ゲームを適用することで最適な利益配分を決定する手法を提案した。

(4) ブロックチェーンの DRM への応用と問題点の検討

ブロックチェーンの DRM の応用が提案されているが、所有権の移転を表現可能なスマートプロパティを適用したとしても不正コピーを防止するためには従来の DRM のフレームワークを用いなければならない。そこでブロックチェーンを契約の表現に適用したスマートコントラクトと従来のコピー防止技術を組み合わせることにより、信頼できる第三者を必要としない DRM の手法を提案した。

3 二次利用を想定した情報カプセルの提案と実装 [目的 2]

3.1 序論

3.1.1 背景

近年、PC の高性能化やネットワークの高速化により、音楽、動画、画像などのデジタルコンテンツの流通が盛んになってきている。デジタルコンテンツの最大の特徴として、無劣化で複製が容易であることが挙げられる。P2P ネットワークなどを介し、無断送信が禁止されている映像や音楽ファイルの不正流通が社会問題となっており、DRM（デジタル著作権管理）といったコピー制御機構の確立が緊急課題となってきている。DRM に関してもまだ改善すべき点が多い。また、TPP（=環太平洋戦略的経済連携協定）合意により著作権侵害の非親告罪化が想定される。著作権侵害が非親告罪化してしまうと、二次創作など同人文化に影響が及び、表現の幅が狭くなってしまふ恐れがある。

また、著作権だけではなく個人情報に関する権利の保護を求める声も高まっており、情報セキュリティの対策が必要不可欠なものとなっている。一方でオープンソース等の配布や改変を認めているという流れもある。従って著作権の所有者とそれを利用するユーザの間には様々な権限に対する要求に柔軟に対応可能なシステムが求められる。^[15]

3.1.2 先行研究

デジタルコンテンツの流通に暗号化されたデジタルコンテンツと、それに対する暗号鍵、著作権情報、管理エージェントなどをパッケージ化した情報カプセルを導入し、このカプセル内の情報カプセルエージェントとアクセス制御エージェントが交渉を行うことにより、コンテンツを管理するという研究がある。しかし、エージェント間で通信を行う際に流通後のコンテンツの位置の把握が困難であるという問題や、利用者などのプライバシーを保護する必要が出てくる。流通における自律管理を目

的としたカプセル化コンテンツの研究がある。これはコンテンツとコンテンツの利用を制御するプログラムをカプセル制御機構としてカプセル化し、コンテンツを利用するソフトの編集機能の有無を判断することで利用制御を行う。しかし、不正利用を防止することを重視している^[3]。

またXMLを用いて権利の記述をすることで、権利の保護や流通の支援を行う研究もされている。こちらは権利情報をXMLに記述して登録することで、XMLの優れた記述解釈と検索性を活用してデータベースとして流通の支援を行う^[11]。

また、我が研究室ではDublin Core Metadata Element SetとDCタームズでの著作権に関する部分の拡張を行うことでコンテンツの権限設定を柔軟に変更することができるシステムの研究を行っている。明確にした利用目的を権限と見なすことでルールを設定し、カプセルエージェントを用いてコンテンツをカプセル化、アクセス制御エージェントでカプセル化コンテンツを管理する。そしてTake-Grantを用いてカプセルエージェント——アクセス制御エージェント間で権限の調停を行う^[13]。これらの研究はどれもユーザがコンテンツを二次利用する場合など、想定されていない利用に対して柔軟な対応はできないと考えられる。

3.1.3 目的

本研究では、Javaプログラムを用いてXMLファイルをベースとしたコンテンツ権限の調停のシステムの提案および実装を試みる。XMLベースでやり取りを行うことで、参照するデータベースとしての役割と同時にユーザの意図を自身で入力可能になるため確実に伝えることが可能となる。コンテンツは想定されていない利用をする必要がある場合がある。その場合あらかじめ利用方法を決めてあると、必要であるにもかかわらずできなくなってしまう。権利者とユーザ間で直接やり取りを行うシステムによって、ユーザの利用目的に対し柔軟な対応が可能とすることを目標とする。

3.2 基礎知識

3.2.1 情報カプセル

情報カプセルは、デジタルコンテンツ自体と、コンテンツの表現に関するもの、コンテンツの利用を制御・制限するもの、コンテンツの利用制御を行う機能の制御をするものなどをカプセル化したものである。コンテンツの制作者によって利用条件が設定され、利用回数や期間、用途など様々な項目でユーザの利用を制御・制限することができる^{[1][2][3][4][9][10][14]}。

3.2.2 XML

XMLとは、文書やデータの意味や構造を記述するためのマークアップ言語の一つ。マークアップ言語とは、「タグ」と呼ばれる特定の文字列で地の文に情報の意味や構造、装飾などを埋め込んでいく言語のことで、XMLはユーザが独自のタグを指定できることから、マークアップ言語を作成するためのメタ言語ともいわれる。

XMLにより統一的な記法を用いながら独自の意味や構造を持ったマークアップ言語を作成することができるため、ソフトウェア間の通信・情報交換に用いるデータ形式や、様々な種類のデータを保存するためのファイルフォーマットなどの定義に使われている。

XML を様々な場面で利用しやすいよう、関連技術の規格も数多く存在する。文書を表示する際の書式や装飾などを指定する XSL、ハイパーリンク機能を実現する XLink/XPointer、XML ベースの言語の仕様を記述するためのスキーマ言語である XML Schema や RELAX、XML をプログラムで利用するための API である DOM や SAX などである。^{[6][11][12][19]}

3.2.3 Dublin Core (ダブリン・コア)

メタデータはデータについてのデータであり、インターネットに分散配置された。データベースの中にどのようなデータがあるかを記述し、検索効率を上げる目的がある。ウェブ上のリソースを記述する共通のメタデータ標準などを開発、促進する組織が多種多様なメタデータを効率的に参照、交換するために必要最低限のメタデータの組み合わせ（メタデータセット）として、Dublin Core Metadata Element Set（以下 DCMES）が開発され、2003年2月には [ISO15836] として国際標準となった。^{[5][13]} DCMES には、中核として 15 個の基本要素が存在する。

3.2.4 関連研究

Bee-gent 東芝研究開発センター知識メディアラボラトリーが開発したエージェントフレームワークで正式名称は Bonding and Encapsulation Enhancement a GENT である。既存アプリケーションをエージェント化するエージェントラッパーと、アプリケーション間の連携手続きを組み込む仲介エージェントの 2 種類で構成される。仲介エージェントはアプリケーションの存在する場所を移動し、エージェントラッパーと情報を交換（対話）しながら連携手続きを実現する。エージェントラッパーはアプリケーションの状態を管理し、必要に応じてアプリケーション処理を起動することで仲介エージェントからの要求に応える。

コンテンツ流通における自律管理を目的としたカプセル化コンテンツ Matryoshka Matryoshka はコンテンツ自身の他にコンテンツ自身に関するもの、コンテンツの表現に関するもの、コンテンツの利用に関するもの、Matryoshka の制御に関するもの大きく分けて四つの各種情報や制御機能を内包する。このシステムの利用条件では、コンテンツ利用の回数、期間、期限、利用者端末制限を行う。利用条件の判断は、編集機能のあるアプリケーションによる処理かどうかで行う。編集機能のないアプリケーションの場合は単純利用であるとして利用条件を発動し、編集アプリケーションの場合は履歴情報や編集前後のデータ差分から判断して適切な利用条件を発動する。^[3]

コンテンツの複合的権利記述による権利保護と流通支援 これはコピーマートモデルに基づく電子著作権取引システムのプロトタイプ構築についての研究である。コピーマートモデルとは著作権者と利用者が契約を介して取引を行う著作権市場の法モデルで、権利者が著作物の提供条件を自由に決定した後、著作物とその利用条件をコピーマートに登録し、利用者は権利者の提示した条件に合意の上で著作物を入手するというものである。このモデルを基に、権利記述登録プログラム、権利記述検索プログラム、権利記述解釈プログラムによって権利情報を XML に記述してデータベース化し、コンテンツの権利保護と流通の支援を目的としている。^[11]

表1 DCMES の 15 個の基本要素

DC 要素名	定義
title (タイトル)	リソースに与えられた名前
creator (作成者)	リソース内容の主たる責任者
subject (主題およびキーワード)	リソースの内容の主題
description (内容記述)	リソースの内容の説明
publisher (公開者)	リソースを提供している主体
contributor (寄与者)	リソースの内容に貢献している者
date (日付)	リソースのライフサイクルにおける出来事に関するとき、もしくは期間
type (資源タイプ)	リソースの内容の種類またはジャンル
format (形式)	リソースの物理的または電子的形式
identifier (資源識別子)	あるコンテキストにおけるリソースへの曖昧のなさの参照
source (出处)	元となったリソースへの参照
language (言語)	リソースの内容を記述する言語
relation (関係)	関連するリソースへの参照
coverage (時間的・空間的対象範囲)	リソースの範囲または領域
rights (権利管理)	リソースの権利に関する情報 知的財産権・著作権・財産権を含む

表2 拡張要素

DC 要素名	定義
request (要求)	拡張タグリソースの利用に関するユーザからの要求情報
purpose (目的)	拡張タグリソースの利用に関するユーザの要求目的
detail (詳細)	拡張タグリソースの利用に関するユーザからの要求目的についての詳細

3.3 提案手法

前述の通り DRM では、二次利用における、ユーザのコンテンツ利用についての柔軟性に問題がある。また、コンテンツの制作者は流通したコンテンツの権限設定を変更することができない。そこで制作者とユーザ間の、コンテンツを二次利用する場合における権利調停の対応を柔軟に行えるシステムを提案する。そのために、まず DCMES の著作権に関する部分とユーザの権限および新たな権限の要求に関する部分を拡張する。

ユーザの権利情報を XML で記述し、情報カプセルがこの XML を読み込むことで、ユーザの利用制限を行う。ユーザが二次利用する申請の流れを図 1 に示す。ユーザがコンテンツを二次創作等に見えるよう申請する場合、情報カプセルで申請内容を記述して XML ファイルとして出力し、ファイル共有を用いて著作者側にアップロードする。著作者はアップロードされた XML を確認後、課金等制限変更の条件を提示する。ユーザが条件を達成した後、著作者が XML の内容を変更することで権限の変更が可能となる。本研究では Java プログラムによって利用権限の申請を XML ファイル形式で出力し、Dropbox 等のファイル共有を用いたアップロードによってコンテンツ著作者とファイルベースでのやり取りを行うことで実装する。

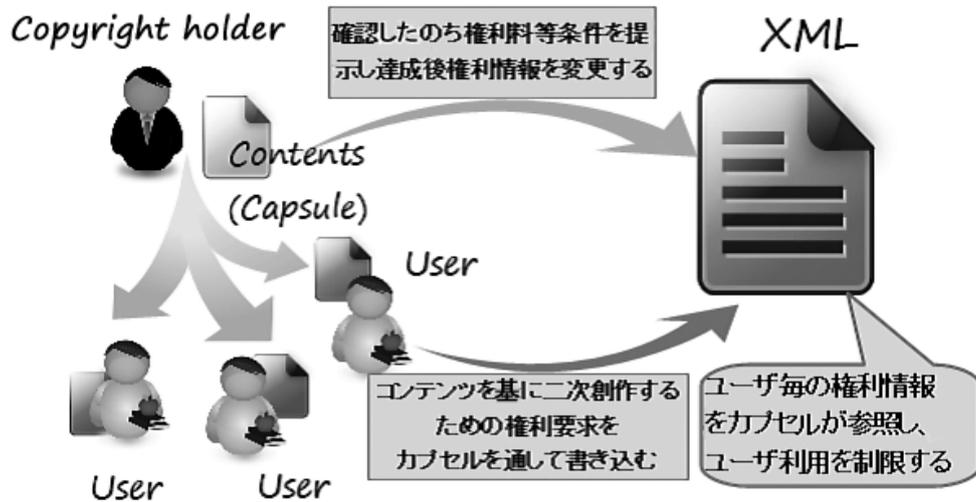


図1 提案システム図

3.3.1 利用制御プログラム

今研究では、二つのプログラムをプロトタイプとして実装した。プログラムソース全体は付録に記述する。

一つ目はユーザの利用を制御するためのプログラム（以下、利用制御プログラム）で、Java パッケージ内に組み込まれている XML ファイル（以下、input.xml）を参照してユーザの権限を通知する。input.xml の例は以下である。



図2 利用制御プログラム実行例

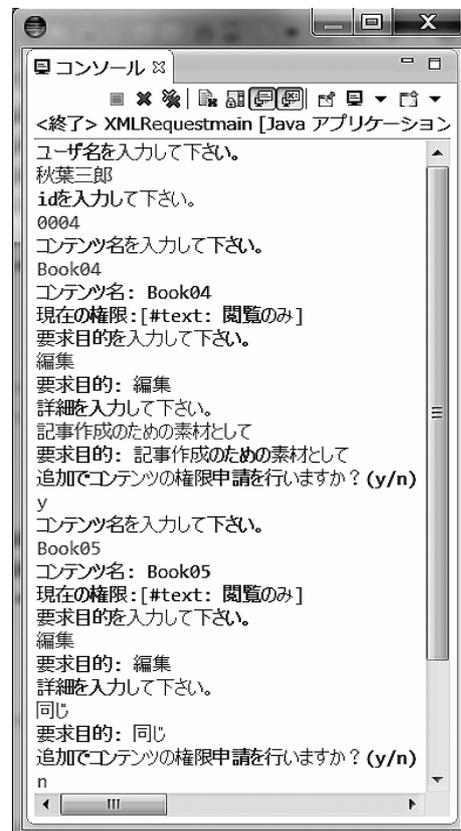


図3 権限要請プログラム実行例

```

<?xml version="1.0" encoding="Shift_JIS"?>
<content title="Picture01">
  <creator>YamadaTarou</creator>
  <users>
    <user name="秋葉一郎" id="0001">
      <rights>View Only</rights>
    </user>
    <user name="芳藤海宏" id="0002">
      <rights>View & Edit</rights>
    </user>
  </users>
  <rights>View Only</rights>
</content>

```

コンテンツを利用する際に利用制御プログラムを起動する。コンテンツ名、ユーザ名、ユーザ id を java.util.Scanner クラスを用いてキーボード入力させることで一致する情報をプロジェクトに含まれる input.xml から参照し、ユーザ固有の権限情報である user タグ内の rights タグ情報と、注意喚起のポップアップを表示する。XML の参照には XML パーサの Xerces を利用している^[16]。実行結果例は図 2 のようになる。ここで dc:rights 等の Dublin Core を用いていないのは、Java プログラムで参照する際に正しく実行されないためである。また見やすいようにインデントしてあるが、input.xml は利用制御プログラムの参照のみで利用するため、実際はインデントをしていない。

3.3.2 権限要請プログラム

二つ目はユーザからの要望を送るプログラム（以下、権限要請プログラム）で、ユーザが持つ利用権限に加え、さらなる権限を要求する場合に著作者などの権利者へ要望として送る XML ファイル（以下、requestID 名.xml）を作成し、共有フォルダへ送る。閲覧のみが許されている場合の requestID 名.xml の例として、以下のようになる。

```

<?xml version="1.0" encoding="Shift_JIS" standalone="no"?>
<contents>
<content title="Book04">
  <dc:creator> 山田太郎 </dc:creator>
  <user id="0004" name="秋葉三郎">
  <dc:right> 閲覧のみ </dc:right>
  <dcterms:request>
  <dcterms:purpose> 編集 </dcterms:purpose>
  <dcterms:detail> 記事作成のための素材として </dcterms:detail>

```

```
</dcterms:request>
</user>
</content>
<content title="Book05">
<dc:creator> 山田太郎 </dc:creator>
<user id="0004" name="秋葉三郎">
<dc:right> 閲覧のみ </dc:right>
<dcterms:request>
<dcterms:purpose> 編集 </dcterms:purpose>
<dcterms:detail> 同じ </dcterms:detail>
</dcterms:request>
</user>
</content>
</contents>
```

今権限要請プログラムではコンテンツのさらなる権限が欲しい場合に起動し、ユーザ名と id、そしてコンテンツ名と利用目的ならびにその詳細を同じく java.util.Scanner クラスを用いてキーボード入力させ、入力に基づき requestID 名.xml を作成する。例として入力がユーザ名「秋葉三郎」id「0004」の場合、「request 秋葉三郎 0004.xml」となる。その後 java.nio.FileChannel クラスのプログラム内に指定してあるディレクトリへコピーされる。実行例は図 3 となる。著作者がこの requestID 名.xml を確認したのちに、ユーザに権限変更条件の提示等手続きを行い、完了後利用制限プログラムで参照する input.xml の内容を書き換えることで権限調停が完了となる。以上二つのプログラムによって、ユーザの意図した二次利用のためのやり取りが可能になる。

3.4 結論

情報社会の現在、デジタルコンテンツでの流通はますます広がっていくのは想像に難くない。これからの著作権管理は著作者の権利が侵害されないよう厳格でなくてはならないのと同時に、ユーザの意図した二次利用が手軽に可能となる柔軟なシステム体系が必要になるであろう。

本稿では、二次利用を想定した場合の権利調停を柔軟に行うことを第一に想定して研究したが、ビッグデータなど大規模データが注目されている現在、セキュリティや大規模データに対する面でもまだ不安が残る。今後の課題としてセキュリティ面、および大規模なデータ量に対する適応、そしてユーザがより使いやすくなるような User Interface の実装が必要であると考えられる。

4 RBAC とハイパーグラフを用いた推論攻撃に対する個人情報保護 [目的 3]

4.1 序論

4.1.1 背景

近年、企業や個人が扱う情報量は増加しており、その情報が漏えいする可能性もまた増加している。情報漏えいを防止するために様々な研究が進められているが、個人情報漏えいが後を絶たない。情報漏えいを防ぐことの本来的な目的は、機密情報全てが漏えいしないようにすることである。そのために研究されている技術の一つがアクセス制御である。アクセス制御はファイルやデータにあらかじめユーザに対するアクセス権限を設定する方法である。アクセス制御により権限のないユーザは直接アクセスできないよう処理がなされている。その中でもロールベースアクセス制御（以下RBAC）は最近開発され、企業などの組織体でのアクセス制御に活用されている。^{[74][75]} RBACは役割によってデータやファイルへのアクセス権が違うので、複数の役割がある企業でのアクセス制御に向いている。しかし、アクセス制御により直接的なアクセスを防ぐだけでは、推論攻撃による機密情報の漏えいを防ぐことができない。^[65] 推論攻撃とは、実行を許可された問い合わせのみを用いて、許可されていない問い合わせの実行結果を推論して得ようとする試みのことをいう。推論によって本来は得ることができない機密情報が外部に流出してしまう可能性が出てきたのである。これにより、複数の情報が集まることにより情報漏えいが起きやすくなってきている。

また、2016年1月より「マイナンバー制度」の運用が開始され、国民一人一人に12桁の番号が割り当てられ、氏名、住所、生年月日などの個人情報を番号で一元管理するようになった。^[66] この制度による個人情報の漏えいが懸念されている。マイナンバー制度において国民一人一人に個人番号カードというICチップ付きのカードが交付される。このICチップに記録されるのは氏名、住所、生年月日、個人番号、本人写真などでプライバシー性の高い情報は記録されない。つまり、個人番号カードからの情報漏えいはほぼ問題ないと考えられる。しかし市役所などの行政内部に悪意を持って情報を流出させる職員などがいないとも限らない。

神奈川大学 木下宏揚研究室 2013年度卒業論文、“マイナンバー制度におけるアクセス制御”ではマイナンバー制度を例にハイパーグラフを用いて推論経路を検出し、制御していた。^[29]

4.1.2 問題点

一般に行政が扱うデータは膨大で、さらにマイナンバー制度では市役所と市役所、または部署と部署の間に膨大なデータが行き来することとなる。一つの部署で扱っていたデータが複数の部署間で扱われることとなると、少量だったデータが膨大になりデータ移行の際に改ざんや情報漏えいが起こる可能性がある。また、利用者には次のような懸念があるとされる。

- 情報を取られたくない
- 追跡されたくない
- 都合の悪い情報は忘れてほしい

つまり個人情報の流出を恐れて情報を残したくないのである。さらに集積・集約された個人情報を基に推論によって、流出元の知人である特定の個人の個人情報が暴かれてしまうといった懸念があ

る。さらに複数のデータが一つに集まった際の処理が複雑になることも考えられる。マイナンバー制度ではマイナンバーに所得情報、納税実績、社会保障などの情報を紐づけて手続きを簡略化することを目的としている。紐づけることによってマイナンバーが漏えいした時にそのものの様々な個人情報の漏えいにつながる危険性は否定できない。具体的には、

- 誰がどのように利用するのかわからない
- 目的不明確な名称により軽々しい扱いを助長してしまう
- 高度な情報収集により情報への勝手な意味づけがされてしまうなどの危険性がある。

4.1.3 目的

本研究ではハイパーグラフによる推論経路分析にロールベースアクセス制御モデルの「役割」という主体を制約条件として付加することで、アクセス権限管理の効率化、推論経路の削減ができると考え、ハイパーグラフによる推論経路分析を主体（役割）と客体の両面から評価するセキュリティモデルを提案する。

4.2 基礎知識

4.2.1 アクセス行列

アクセス行列とは主体（subject）と客体（object）の関係を表した行列のことで、主体と客体の関係には R（Read：読み書き可能）、W（Write：書き込み可能）、RW（Read+Write：読み書き可能）、 ϕ （Phi：読み書き不可）の4種類の権限がある。

/	S_1	S_2
O_1	ϕ	W
O_2	R	R

図4 アクセス行列

主体（subject） 主体とはネットワークやデータベース内で管理されている客体にアクセスする行為者であり、ユーザに相当する。

- 名前……主体の名前
- 競合……管理している主体のコミュニティの情報
- 階層……コミュニティで指定されたセキュリティレベル
- 役割……客体の権限を決定する役割

客体（object） 客体はネットワークやデータベース内で管理されている情報であり、ファイルに相当する。

- 名前……客体の名前
- 競合……管理している主体のコミュニティの情報
- 階層……コミュニティで指定されたセキュリティレベル
- 所有……管理している主体の情報
- プライベート……管理している主体の情報

コミュニティ (Community) コミュニティとは、コミュニティの属性、コミュニティに属する主体、および、コミュニティが管理する客体とその属性の集まりから成る社会システムに相当する。コミュニティ (Community) 同士には利害関係があり、管理している主体には組織的に階層レベルや役割を割り振られる。コミュニティにも様々な種類があるが、インターネット上のコミュニティを考えると主体の役割や利害関係、あるいはプライベートな情報（個人情報）が複雑に絡み合っている。現在、求められているのはこのように複雑に絡み合ったコミュニティにおいて実現するセキュリティモデルである。それが実現されたのが Community Based Access Control Model である。

4.2.2 Covert Channel

間接情報フロー Covert Channel とはアクセス行列において、本来客体 (Object: データやそれを含む情報) に直接アクセスする権限 (Permission: アクセス権) がない主体 (Subject: 利用者、ユーザ) なのにもかかわらず、アクセス権を持つ第三者の力を借りて間接的にその客体を操作できるようになってしまうことである。その時に発生する客体に対しての主体へのアクセス権限が矛盾した不正な経路を Covert Channel という。またこれを間接情報フロー^[70]と呼ぶ。以下の流れが例である。初期状態 S_2 は直接 O_1 の情報を読み込むことができないが以下の流れで読むことができてしまう。

1. S_1 (Subject) が O_1 (Object) を読み込む
2. S_1 が O_1 で読み込んだ情報を O_2 (Object) に書き込む
3. S_2 (Subject) が O_2 を読み込む
4. 発生した Covert Channel より読めないはずの O_1 の情報を S_2 が読める

このような流れで不正な情報流出が発生してしまうためアクセス制御を行う推論エンジンとしては、できる限りこれが発生するのを防ぐ検出と訂正を的確に行えるようにするのが、情報フィルタに必要とされる機能である。

実際に発生する Covert Channel 不正な情報経路である Covert Channel を全て塞いでしまえば安全なシステムを構築することができるように見えるが、単独では隠れチャンネル (Covert Channel) が存在しないようなコンピュータでもネットワークに接続されたコンピュータ群が協調することによって、隠れチャンネルを構成できてしまう。つまり、単独では安全なコンピュータでも、それがネットワークを構成すると安全ではなくなるような状況が簡単に存在し得るのである。このようなネットワーク構成機能の問題点が Covert Channel で利用される。以下に例が挙げられる。

- 会社の機密データを社外へ持ち出したり、社外の人間 (社外の PC) でも見られるようにする
- mixi 等の SNS の個人データが掲示板やブログ等不特定多数へ流出
- スパイウェア等、個人 PC から情報を持ち出すため

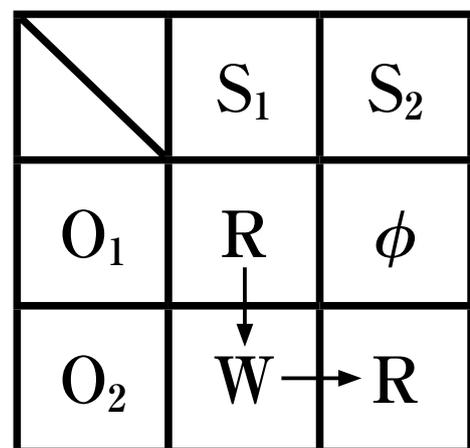


図5 Covert Channel の例

にこれを用いて通信を行い、検知を困難にさせる

WWW 等、不特定大多数が利用するネットワークでは意図しなくても Covert Channel が発生してしまう恐れがあるのでそういった情報網では比較的安易に情報漏洩が起こりうる。このように Covert Channel は今のネットワーク社会にとって情報を安易に流出させてしまう存在なのである。

Community Based Access Control Model Community Based Access Control Model は、Covert Channel の制御を実現するためのセキュリティモデルの一つである。^[68] Access Control Agent System がこのモデルには組み込まれていて、コミュニティを用いた Covert Channel 分析が行える。まずユーザ数とファイル数を抑制し、整理するために共通する属性を持った小規模なユーザの集合を Community と定義する。各コミュニティではそれぞれ内部で Covert Channel 分析を行い、Covert Channel がおきないように制御する。外部コミュニティと通信した場合の Covert Channel 分析は自コミュニティ、アクセス要求者、その要求について全て分析する。Covert Channel 分析は、Subject や Object の関係を以下のようなアクセス行列で表現し、Covert Channel の検出をする。属性はアクセス制御や Covert Channel との関連性から競合、所有、階層、役割、プライベートの五つを使用する。このとき、アクセス行列から図 4 のような 2×2 行列のパターンを全て取り出し、Covert Channel 分析を行う。このやり方により、 2×2 行列全てのパターンだけでなくそれらを組み合わせることで 3×3 やそれ以上の場合も Covert Channel を全て検出することができる。このようにこのモデルは Covert Channel を防ぐのに元々適したモデルであるので、この推論機能をベースとして改良し、新たな処理のアクセス制御を行う。

セキュリティモデル セキュリティモデルは、アクセス制御システムを構築する上で、セキュリティポリシーを具体的な論理的形式で表現したものである。^[68] そこには制御したいサービスや組織構造が反映される。最も単純な型は、permission (read, write, \neg read, \neg write) であり、Subject (主体)、Object (客体) を含めた三つでアクセストリプルと呼び、それをシステムで如何に扱うかによってアクセス制御が行われる。従来のセキュリティモデルには様々な種類があり、使用される状況に応じて使い分けたり、複数使用する等している。

情報フィルタ 情報フィルタとは Covert Channel 検出時にその Covert Channel がなくなるように特定の権限を変更することである。情報フィルタには 4 種類の方法があり、それぞれ一長一短がある。3 種類はフロー経路の権限を禁止して遮断するのに対し、Read 権限を許可する方法は情報共有の拡大の意味を持つ。以前は読めなかった客体が修正により普通に読めるようになれば、不正経路ではなくなるので Covert Channel 自体はなくなることができる。情報フィルタの具体的な処理を以下にまとめていく。図のように Covert Channel が発生して検出された場合、以下、図 6 の(a)(b)(c)(d)のいずれかを適用すれば Covert Channel が解消される。

- (S_1, O_1) の READ 権限を削除
- (S_1, O_2) の WRITE 権限を削除
- (S_2, O_1) に READ 権限を添付

- (S_2, O_2) の READ 権限を削除上記のどの情報フィルタを選択するかは各コミュニティのセキュリティポリシーや主体のアクセス履歴、ユーザがどういう方針で処理するかを定めるユーザポリシーを考慮して決定するが、(a)から(d)のどの場合でも Covert Channel は訂正できる。

\	S_1	S_2
O_1	ϕ	R
O_2	ϕ	RW

(a)

\	S_1	S_2
O_1	ϕ	ϕ
O_2	R	RW

(b)

\	S_1	S_2
O_1	ϕ	R
O_2	R	W

(c)

\	S_1	S_2
O_1	R	R
O_2	R	RW

(d)

図6 情報フィルタ

4.2.3 推論による情報漏えい

一つ一つの情報それ自体は秘密情報でなかったとしても、それらが複数集まり何らかの推論を施すことによって、秘密情報を抽出できてしまうことがある。そのような攻撃を推論攻撃と呼ぶ。推論攻撃への対策はデータベースのセキュリティ課題として研究されてきた。推論攻撃の一種として、情報間の統計的な関連に注目した研究がある^[69]。データベースに蓄積されているデータは、それぞれが無関係に独立に存在しているわけではなく、統計的あるいは意味的に関連している場合が多い。例えば、ある発言者が同じ地名、もしくは駅名などの単語を頻繁に発言している場合、この地名や駅名と発言者との間に何らかの関連があることが統計的に分析できてしまう。推論は統計的手法以外にも様々なものがあり、それら全ての推論解析攻撃に対抗するためには膨大な情報群とその間にある推論関係を常に監視し、何らかの問題を未然に検知して警告するようなシステムが必須である。しかし、そのためには情報間の推論的依存関係を記述するモデルが必要である。どのような推論手法であっても、いくつかの情報からある情報を導くことに変わりはない。その依存関係をモデル化できれば、推論手法によらない対策を考えることができるかもしれない。図7はあるオブジェクト集合におけるオブジェクト間の依存関係を洗い出してリスト化したものである。

推論元オブジェクト	推論	導出オブジェクト
O_1, O_2, O_3	\Rightarrow	O_4
O_4, O_6	\Rightarrow	O_5
O_3, O_6	\Rightarrow	O_8
O_6, O_8	\Rightarrow	O_7
O_4	\Rightarrow	O_6

図7 object間の依存関係リスト

4.2.4 有向グラフと頂点着色によるモデル化

グラフの頂点着色 グラフ $G=(V, E)$ の頂点着色とは、グラフの頂点に色を塗ることである。すなわち、ある色集合 C と写像 $c:V \rightarrow C$ を与えることである。特に $(v, w) \in E$ であるような、どの2頂点 $(v, w) \in V$ についても $c(v) \neq c(w)$ となるとき、 c を頂点彩色と呼ぶ。色数 $|C|$ が制限されている時に頂点彩色可能かどうかを判定する問題や、可能ならばその彩色を求める問題はグラフ理論において重要な問題として研究されており、様々なアルゴリズムが考案されている。一般に、ある条件 P を満たす頂点着色を P 着色と呼ぶことにする。頂点着色は次のように一般化できる。グラフの頂点にあらかじめ色のリストが与えられているとする。すなわち、写像 $L:V \rightarrow \mathcal{2}^C$ が与えられているとする。このとき、 G の頂点着色 c で全ての頂点 $v \in V$ に対して $c(v) \in L(v)$ を満たすものを G のリスト着色と呼ぶ。任意の頂点 $v \in V$ に対して $L(v) = C$ ならば通常の頂点着色である。 c が頂点彩色のときは、特にリスト彩色と呼ぶ。また、 c が P 着色のときは P リスト着色と呼ぶことにする。各頂点 v における色リストのサイズ $|L(v)|$ が制限されている時にリスト彩色可能かどうかを判定する問題や、可能ならばその着色を求める問題が通常の頂点彩色と同様に広く研究されており、様々なアルゴリズムが考案されている。

推論による頂点着色のグラフ表現 まず、オブジェクト間の依存関係リストを次のようにグラフ化する。頂点集合 V をオブジェクト集合とし、依存関係リスト上で $O_{i1} \dots O_{ik} \in V$ から $O_j \in V$ が導出可能ならば有向辺 $(O_{i1}, O_j) \dots (O_{ik}, O_j)$ を描く。さらにそのグラフに対する色リストを用いて ACL を次のように表現する。ただしここでは議論を簡単にするために ACL において read 可能か否かのみに着目

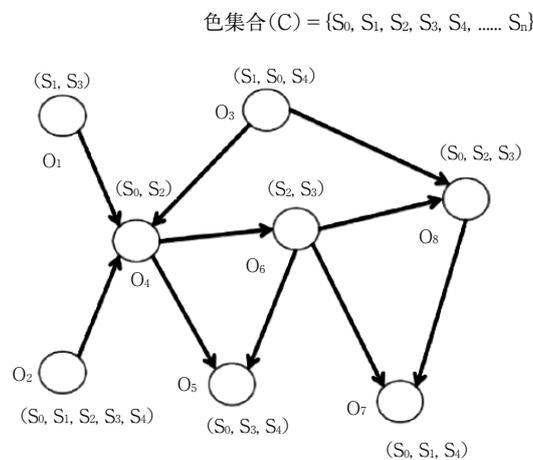


図8 推論的依存関係と ACL の有向グラフ表現

する。まず、色集合 C をサブジェクト集合とする。そして、ACL 上で、あるサブジェクト $S_i \in C$ があるオブジェクト $O_j \in V$ を read 可能ならば、 $S_i \in L(O_j)$ とし O_j の色リストに S_i を加える。例として、図 7 をグラフで表現し、ある ACL に従って色リストを与えたものを図 8 に示す。

サブジェクト S_i がオブジェクト O_j を read したとき頂点 O_j に色 S_i を塗るとしよう。例えば、図 8 において O_1, O_2, O_3 が全て S_1 で塗られたとする。このとき、推論によって O_4 が導出可能なことを考えれば O_1, O_2, O_3 が全て S_1 で塗られた時点で O_4 も S_1 で塗るべきである。しかしその一方で、 $S_1 \notin L(O_4)$ 、すなわち、ACL 上では S_1 は O_4 を read できないことになっているので、これは推論による情報漏えいを意味している。このとき、その頂点着色はリスト着色の定義にも反していることに注目すると、推論による情報漏えいに関する安全性を次のように定義できる。定義：ある P 着色がリスト着色ならばその着色は推論に対して安全であるという。ここで、 $P = \text{「任意の頂点 } v \text{ に対して、} v \text{ を終点とする全ての有向辺 } (u_1, v) \dots (u_k, v) \text{ の始点 } u_i \text{ が同一色で塗られているならば、} c(v) = c(u_i) \text{ でなければならない」}$ とする。

有向グラフ表現の問題点 有向グラフでは表現不可能な依存関係リストが存在する。例えば図 9 において O_1, O_2 から、 O_4 が導出でき、 O_1, O_3 から O_4 が導出できる。これを有向グラフで表現しようとすると O_1, O_2, O_3 から O_4 へ有向辺を描くことになり、しかしそれでは O_1, O_2, O_3 から O_4 が導出できるという意味になってしまう。

推論元オブジェクト	推論	導出オブジェクト
O_1, O_2	\rightarrow	O_4
O_1, O_3	\rightarrow	O_4
O_1, O_2, O_4	\rightarrow	O_5
O_1, O_3, O_5	\rightarrow	O_2

図 9 有向グラフでは表現不可能なリスト

4.2.5 ハイパーグラフ

ハイパーグラフの定義 グラフにおいて辺とは 2 頂点对のことであった。^[67] これは辺は 2 個の頂点から

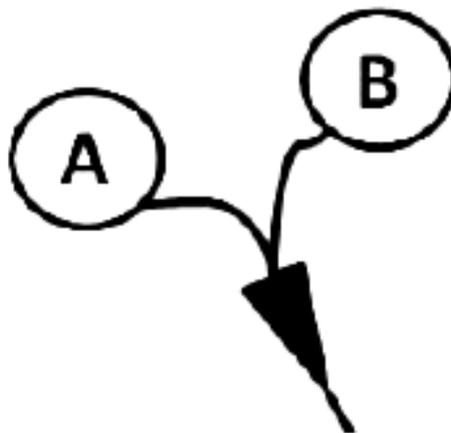


図 10 有向ハイパーグラフ

成ることを意味する。この個数制限を自由にすることで一般化したものがハイパーグラフである。ハイパーグラフは $H=(V, E)$ と記述される。ここで V は頂点集合 $E \subseteq 2^V$ は V の部分集合である。

有向ハイパーグラフの定義 有向ハイパーグラフ $H=(V, E)$ は頂点集合 V と有向辺の集合 E から構成される。ここで有向辺とは空ではない互いに素の V の二つの部分集合 S, T の順序対 (S, T) である。

4.2.6 RBAC

RBAC はユーザの組織内における一定の権限や責任を伴う業務上の役割に応じてアクセス権限を細かく分割し、アクセス制御を行う。^[71] ユーザは、直接ファイルからアクセスを許可してもらわずに、代わりに特定のロールに所属する。

4.2.7 マイナンバー制度

マイナンバー制度について マイナンバー制度とは国民一人一人に個人番号と呼ばれる固有の番号を割り当て、諸手続きの簡略化を目的とした制度である。アメリカなどでは既にこの制度が適用されており、医療、介護、年金などの社会保障などの分野で利用されている。しかし、なりすましなどによって番号が売買されており、主にネット犯罪が横行している。政府はこのなりすまし犯罪に対策を練っているが、解決には至っていない。

日本におけるマイナンバー制度 日本でのマイナンバー制度では、国民一人一人に個人番号カードが交付される。この個人番号カードにはICチップが搭載されており、このICチップには氏名、住所、生年月日、性別、個人番号、本人写真が記録される。しかしながら一方で、プライバシー性の高い情報（地方税関係情報、年金給付関係情報などの特定個人情報）は記録されない。他には総務省が

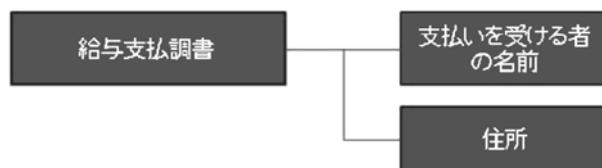


図 11 給与支払調書に格納されている情報

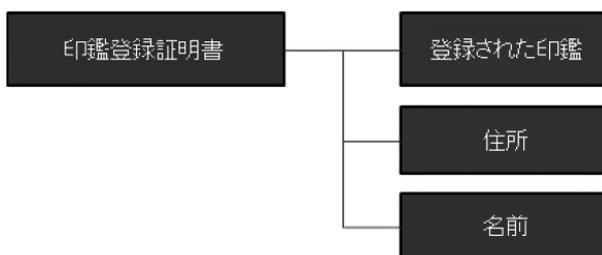


図 12 印鑑登録証明書に格納されている情報

定めた公的個人認証に係る電子証明書など、市町村が条例で定めた事項などが記載される。この個人番号カードから流出する可能性があるのは基本的な4情報などである。

マイナンバーにおける推論 マイナンバー制度では様々な機関にまたがり様々な情報がやり取りされる。さらにその情報にはマイナンバーを通じて多くの情報が紐づけられている。第2章で推論による情報漏えいについて示したように、多いとは言えない情報から推論によって情報漏えいが起きてしまう可能性がある。具体的な例を次に示す。

- 企業が自治体に提出する給与所得支払調書には次のような情報が格納されているとする。
- 市役所が扱う印鑑登録証明書には次のような情報が格納されているとする。

これらの情報について推論を行うと、本来なら登録された印鑑は見られないはずだが、住所と名前からの推論によって登録された印鑑が見えてしまう可能性がある。

4.3 提案モデル

4.3.1 セキュリティモデルの定義

本研究で使用するセキュリティモデルを定義する。

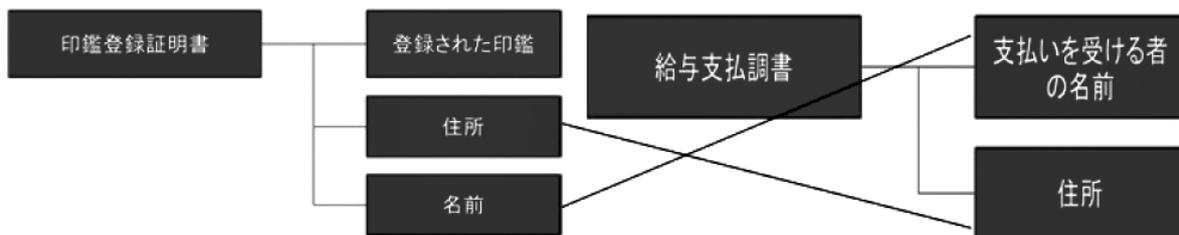


図13 給与支払調書と印鑑登録証明書の結合

モデル要素

Role……役職（役割）

Object……マイナンバー制度で扱う情報

Operation……RW（Read+Write）, Read, Write, ϕ （読み書き不可）

モデルの機能

RBAC……アクセス権を役割によって割り当てる

ハイパーグラフ……Object間の依存関係を効率良く表す

Covert Channel分析……アクセス行列において、客体に対しての主体へのアクセス権限が矛盾した、不正な経路を検出する

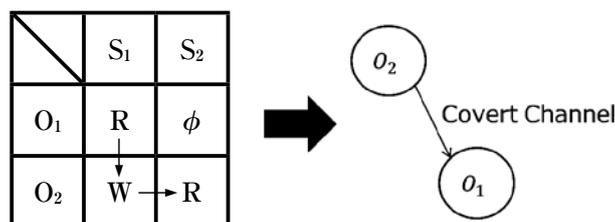


図14 Covert Channelの分析例

分析アルゴリズム

1. Object がどの書類に記載されているか、また、Object の内容が同一かで Object 間の依存関係から Object のハイパーグラフを作成する。
2. 1 のハイパーグラフのノードにアクセス行列から Role を割り当てる。
3. Covert Channel 分析をし、ハイパーグラフに付加する。
4. 下の推論条件を用いてハイパーグラフに推論可能な経路があるか分析する。

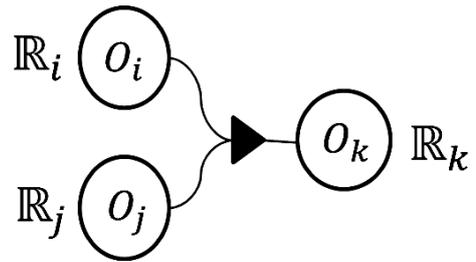


図 15 推論例

Role を R_i 、Role の集合を R とし、図 15 の $(O_i, O_j) \rightarrow (O_k)$ の場合、

(a) 推論可能の条件

$$R_i \in R_i \wedge R_j \in R_j \wedge R_i \notin R_k$$

(b) 推論できない条件

$$R_i \in R_i \wedge R_i \notin R_j \wedge R_i \notin R_k$$

または、

$$R_i \notin R_i \wedge R_i \in R_j \wedge R_i \notin R_k$$

(c) 推論ではない条件

$$(R_i \in R_i \vee R_i \in R_j) \wedge R_i \in R_k$$

4.3.2 マイナンバー制度における分析例

マイナンバー制度におけるセキュリティモデルを設計し、分析をする。

Role、Object を以下に示し、アクセス行列を図 16 に示す。マイナンバー制度における役職

給与支払調書	出生届	印鑑登録書	法定調書
O ₁ …被支払者名	O ₄ …出生者住所	O ₇ …登録印鑑	O ₁₀ …提出者氏名
O ₂ …被支払者住所	O ₅ …出生者生年月日	O ₈ …氏名	O ₁₁ …提出者住所
O ₃ …被支払者生年月日	O ₆ …出生病院	O ₉ …住所	O ₁₂ …提出者電話番号

	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇
O ₁	R	R	RW	φ	φ	φ	φ
O ₂	R	R	RW	φ	φ	φ	φ
O ₃	RW	R	RW	φ	φ	φ	φ
O ₄	R	φ	φ	φ	R	φ	RW
O ₅	R	φ	φ	φ	R	φ	RW
O ₆	R	φ	φ	φ	R	φ	RW
O ₇	R	φ	φ	φ	R	RW	φ
O ₈	R	φ	φ	φ	R	RW	φ
O ₉	R	φ	φ	φ	R	RW	φ
O ₁₀	R	R	φ	RW	φ	φ	φ
O ₁₁	R	R	φ	RW	φ	φ	φ
O ₁₂	R	R	φ	RW	φ	φ	φ

図 16 マイナンバー制度におけるアクセス行列

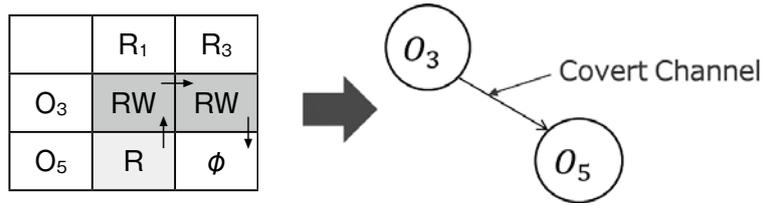


図 17 Covert Channel 分析

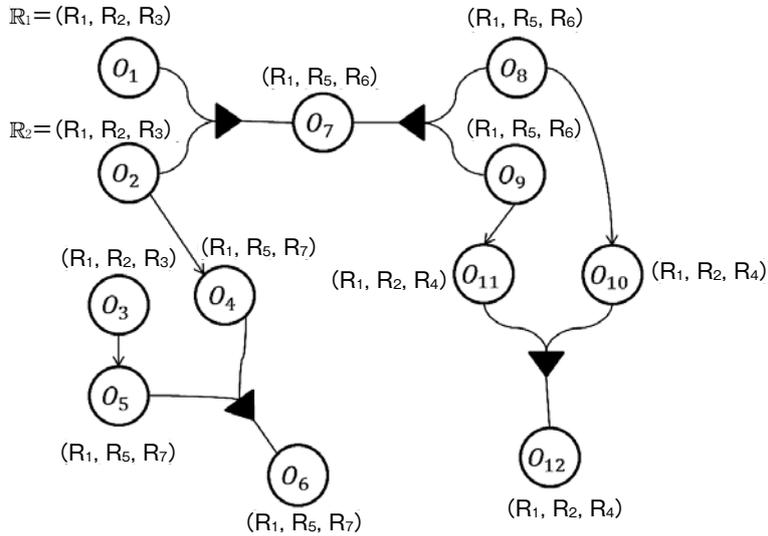


図 18 マイナンバー制度におけるハイパーグラフ

^[72]
(役割) は

- R₁：市長
- R₂：法人課税課課長
- R₃：法人課税課特別徴収担当
- R₄：法人課税課法人市民税担当
- R₅：戸籍課課長
- R₆：戸籍課戸籍担当
- R₇：戸籍課登録担当

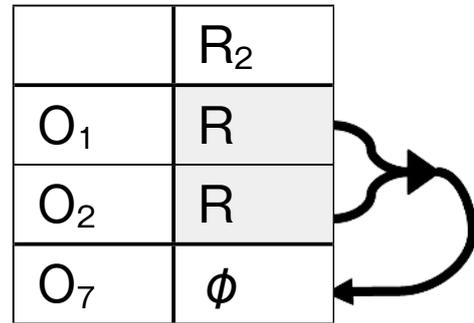


図 19 推論による Covert Channel

マイナンバー制度で扱う書類を要素ごとに Object を割り当てる。

分析アルゴリズムを用いて、マイナンバー制度でのハイパーグラフを作成する。マイナンバー制度で扱う書類から Object の依存関係を表すハイパーグラフができる。それに Role を割り当てる。

また、図 16 では Covert Channel が発生しているため、図 17 のように Covert Channel を考慮する。よって Covert Channel を考慮したハイパーグラフは図 18 になる。

分析例 1 (O₁, O₂)→(O₇) について推論条件を用いて分析する。

(O₁, O₂)→(O₇) は推論条件より R₂ が図 19 のような推論による Covert Channel が発生していることが分かる。R₃ も同様である。推論が可能であるため、推論経路である。

分析例 2 $(O_3, O_4) \rightarrow (O_6)$ について推論条件を用いて分析する。

Role を割り当てない場合、 $(O_3) \rightarrow (O_5)$ であるから $(O_4, O_5) \rightarrow (O_6)$ と同様に $(O_3, O_4) \rightarrow (O_6)$ が推論できてしまう。しかし、Role を割り当てると (O_6) にアクセス権がない (R_2, R_3) は (O_4) を知ることはできない。よって $(O_3, O_4) \rightarrow (O_6)$ は推論不可であるから推論経路ではない。

4.4 結論

本論文では、ハイパーグラフによる推論経路分析にロールベースアクセス制御モデルの「役割」という主体を制約条件として付加することで、アクセス権管理の効率化、推論経路の削減ができると考え、ハイパーグラフによる推論経路分析を主体（役割）と客体の両面から評価するセキュリティモデルを提案した。マイナンバー制度での分析例では、推論による Covert Channel を検出することができ、Role を割り当てることで推論経路を減らすことができた。本手法では、客体の数が膨大である場合や一つの客体を推論する際に三つ以上の客体が必要な場合など客体の数や紐付けのパターンによって結果が異なる可能性がある。また、推論経路検出の効率化をどのように評価するか、推論経路検出の自動化、推論経路検出後の制御が今後の課題である。

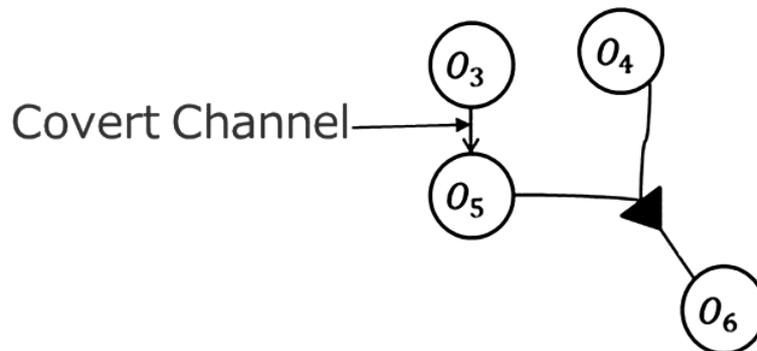


図 20 Role がない場合

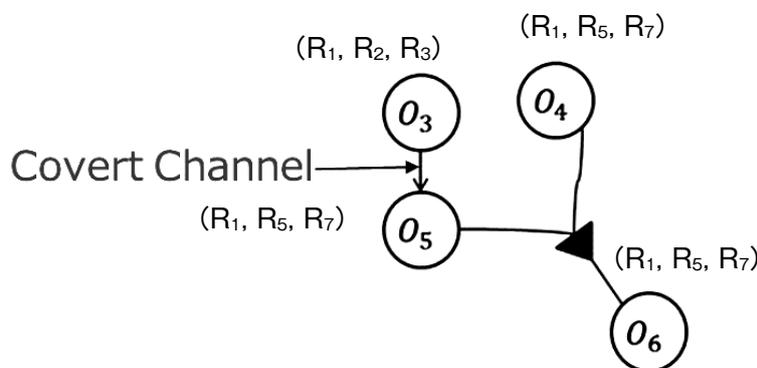


図 21 Role がある場合

5 ハフ変換を用いた電子透かし [目的 2]

5.1 序論

5.1.1 背景

インターネットの普及に伴い、画像や音楽などのデジタルコンテンツが日常生活においてなくてはならないものになった。これらのデジタルコンテンツは簡単に複製が可能^[33]なために在庫切れや絶版がない上、複製元の品質まで同じにできる^[32]ので劣化がない。他にも電子媒体の特性上により場所を取らない、紙といった資源を消費しない、など多くの利点がある。しかし、一方で前述したようにマルチメディアの需要の増大化に伴い、映像ソフトや音楽ソフトのデジタル化については不正コピーの問題がクローズアップされてきた。結果として、著作権者の意図に反してインターネットを介したデジタルコンテンツの第二、第三の無許可コピーが流布する結果を招きつつある。これは著作権者が正当な利益を得る妨げにもなると考えられる。現在、独自のデジタルコンテンツを違法コピーのはびこる情報ネット上に公開するには著作権の保護および管理が必須となりつつあり、これらを防ぐ技術として電子透かしが用いられる。この埋め込まれた透かし情報には著作権者の固有の識別符号があって所有権の所在を主張しようと企図されていたり、本来のコンテンツの改ざん場所を検出するために用いられる。つまり不正にコピーする者への牽制と防止を担っている。

その必要性から、昨今まで電子透かしを使った様々な提案がされてきている。それらの研究目的として電子透かしの耐性の向上やコンテンツの劣化の軽減等が挙げられる^{[33][35]}。しかし、画像の分類として線画を対象とした電子透かしの研究は多くない。要因の一つとして透かし情報を埋め込む対象が画像全体の場合が少なくないためだと思われる。線画だと空白部分が多いため、画像全体に埋め込んだ際に透かし情報が目立ちやすい事がある。また、埋め込み対象が画像全体の場合だと画像自体への影響も考えられる。線画の先行研究として、P型フーリエ記述子を用いて耐性を目的とした線画の電子透かし^[35]の方法がある。

5.1.2 問題点

線画への電子透かしが少なく、電子透かしの埋め込み対象として画像全体や輪郭部分全てに埋め込みを行い、画像の劣化を招いている。

5.1.3 目的

目的として元画像から輪郭部分を一部を抽出し、その箇所に透かし情報を埋め込む。そして、線画に対応できて画像の劣化を軽減する有効性を検証したい。本研究では、従来の研究^[35]と比べてある程度自動で埋め込み箇所が簡単に決定できるハフ変換を用いて輪郭部分の中から直線を検出し、そこに周波数変換での埋め込みを行う。そして、この提案手法の性能と画質の評価を行う。ハフ変換の詳細については節 5.2.2 で記述する。

5.2 基礎知識

5.2.1 電子透かし

電子透かしとは、デジタルコンテンツに他の情報を埋め込む技術のことである。周波数を用いて埋め込んだ電子透かしのイメージを図 22 に示す。

電子透かしは情報を埋め込んだ上でコンテンツ自体の劣化を防がないといけない。また、意図的に透かし情報の消去や改ざんをできないようにさせなければいけない。これらを考えて、電子透かしを利用するとしたら、人間の感覚を利用して認識できないように埋め込む、デジタルコンテンツにある冗長性を利用し人間が知覚できない部分に情報を埋め込む、という二つの方法が考えられる。また、^{[33][34][36][37]}目に見える形で著作権情報を明示する電子透かしもある。電子透かしは以下の分類に分けることができる。



図 22 原画像と埋め込み後の画像

- 不可視・高耐性型 埋め込み可能な情報量が少ないという欠点があるが、加工・圧縮などの処理に対して、耐性が高く、電子透かし情報の除去が困難であるという利点がある。
- 不可視・低耐性型 耐性を重視せずに、多くのデータを埋め込みたい場合に用いる。耐性が必要ないため、画質への影響や劣化も少なくすることができる。コンテンツの改ざんの発見にも利用することができる。
- 可視・可逆型 所有権を明示的にコンテンツ上に表示する電子透かしである。透かしの埋め込んだコンテンツの再利用ができるように、電子透かしを除去することができる。電子透かしを除去するには、特別なプログラムと鍵が必要である。これにより、コンテンツの所有者は、安心してコンテンツをインターネット上で公開したりすることができる。
- 可視・非可逆型 所定のコンテンツの所有権をはっきり利用者に示す際に用いる。この型は埋め込んだ透かしデータは取り除くことができない。不正な再利用や販売を防止する目的を持ち、コンテンツに所有者の名前やロゴマークを付加する。^{[33][34]}
- 埋め込み方式による分類 埋め込み方式は周波数および画素に埋め込むものがある。^[34]
- 画素に埋め込む方式 画像の画素にビットを埋め込む方式。ウェーブレット変換で求めた画像に用いられている。^[33]

- 周波数に埋め込む方式 画像の周波数に埋め込む方式。DCT で周波数領域を求めて埋め込むものが多い。^{[41][42]}

5.2.2 ハフ変換

特殊抽出法の一つであり、画像の特徴的な点を多く通過するものを判定する。それによって元画像に合った線を抽出する。最初にエッジ検出と呼ばれる色調の変化を特定するアルゴリズムを使い、局所的なピクセルの最大値およびそれと同等の値を判別する。そして、判別された複数の点から以下の式を用いて極座標にし、重なり合った点から直線を抽出する。^[38]

$$\rho = x.\cos(\theta) + y.\sin(\theta) \quad (1)$$

$$\theta \text{ (度単位) の範囲 } -900 \leq e < 900 \quad (2)$$

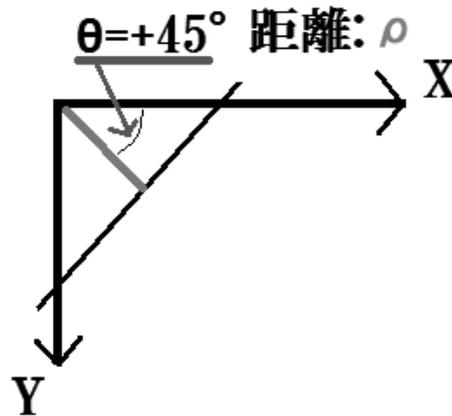


図 23 イメージ図

色調の変化を判別するため、グラデーション画像にハフ変換を行うと図 24 のようになる。

5.2.3 符号化技術

電子透かしを行う際に画像を符号化し、そこに情報を埋め込み、再び画像に変換することで電子透かしとなる。以下にその符号化技術を挙げる。

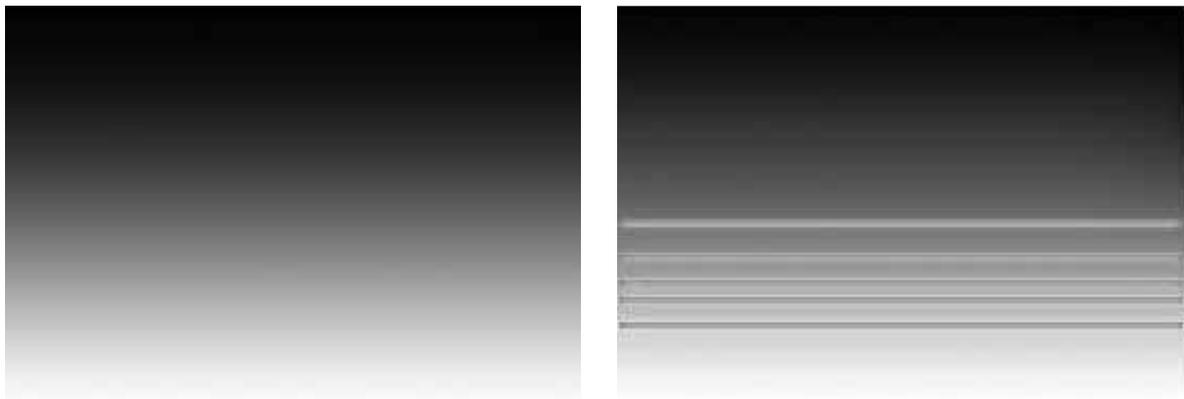


図 24 元画像と変換後の画像

フーリエ変換 ある変数の関数をその変数に共役な変数の関数に変換する方法である。光、音や画像を周波数の関数として表したり、位置の関数としての物体を波数の関数としての回折図形に変換したりするときに使われる。^{[33][34][37]}

離散フーリエ変換 コンピュータ上でフーリエ変換を行う場合、離散フーリエ変換（DFT）と呼ばれる変換を行う。離散変換とは、入力値と出力値が離散サンプルで、コンピュータ操作がしやすい変換のことである。この形式の変換を使用する主な理由は二つある。一つ目に DFT の入力と出力は両方が離散であり、コンピュータで操作しやすい点、二つ目に DFT を計算するために、高速フーリエ変換（FFT）として知られるアルゴリズムがある点である。^{[33][34][37]}

P 型フーリエ記述子 ^[35] 従来研究より P 型フーリエ記述子の主な性質をまとめると以下のようになる。

• (1) 再生曲線の安定性

P 型フーリエ記述子による再生曲線は原曲線とほぼ一致する。つまり、原曲線が閉（開）曲線なら常に再生曲線は閉（開）曲線となる。また、開曲線の場合に、曲線の端点の座標が保存される。

• (2) 幾何学操作に対する P 表現の不変性

P 型フーリエ記述子は P 表現の定義から明らかに並行移動や拡大・縮小に関しては、不変である。また、回転や裏返し、始点位置の変更（閉曲線の場合）等の操作に関しても簡単な関係が成り立っている。

• (3) 原曲線に対する近似の視覚的良さ

再生曲線は原曲線の近似となっていて、P 型フーリエ記述子の低域部分・高域部分はそれぞれ曲線の概形・細部に対応している。

(3)の性質より、P 型フーリエ記述子の中・高域部分に電子透かしの埋め込みを行っても曲線の形に大きな影響を与えないことが分かる。

離散コサイン変換 離散コサイン変換（Discrete Cosine Transform：以下 DCT）とは、画像の情報圧縮符号化に広く用いられている技術である。画像や音声などをサンプリングして離散的な信号に変換し、離散コサイン変換を行った後に符号化を行うことで、元の信号の大部分を損ねずにデータの容量を減らすことができる。JPEG などの画像圧縮技術や AAC、MP3 などの音声圧縮技術において利用されている。DCT では、画像の画素値を、余弦関数を規定とする係数、いわゆる DCT 係数に変換する。これにより、画像領域の値を周波数領域の値に変換することができる。また、DCT 係数から画素値に戻すことを逆離散コサイン変換（Inverse Discrete Cosine Transform：以下 IDCT）という。画像・周波数領域の変換を図 25 に示す。離散コサイン変換は変換後の信号の周波数成分が低周波数領域に集中することが特徴である。データ圧縮に離散コサイン変換を利用する場合、変換後の信号を量子化し符号化する際に、情報の集中していない領域に対して少ない符号化ビットを割り当てるか、または 0 で近似し切り捨てることで、データの容量を減らすことが可能である。^{[33][34][37][41][42][43]} また、本研究では離散コサイン変換を用いる。

計算式は MATLAB で使われている以下の式を用いる。

$$y(k) = w(k) \sum_{n=1}^N x(n) \cos\left(\frac{\pi}{2N} (2n-1)(k-1)\right), \quad k = 1, 2, \dots, N, \quad (3)$$

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leq k \leq N \end{cases} \quad (4)$$

ウェーブレット変換 ウェーブレット変換とは、ウェーブレット関数を平行移動と伸縮の操作を施したものと畳み込みで定義される。ウェーブレット変換は様々な種類があるが、電子透かしにはハール関数がしばしば利用される。ハール関数はウェーブレット変換の中でも最も簡単な関数であり、データの圧縮、伝送という除去などによく利用される。特徴として、対称性、直交性を持つ。そして、計算式が簡単なため計算時間を短くすることができる。^{[33][34][37][44]}しかし、今回ハフ変換を用いた直線の電子透かしにはオクターブ分割を使用するウェーブレット変換は適していないと思われる。

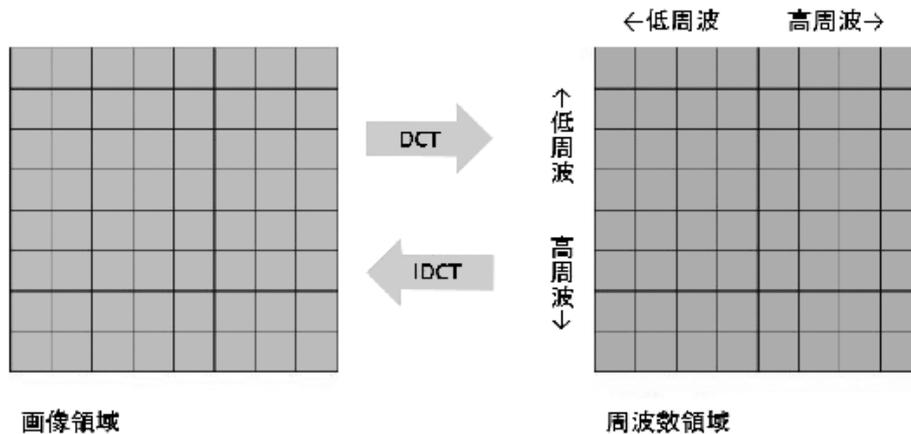


図 25 画像領域から周波数領域への変換

5.2.4 マスキング効果

マスキング効果とは、ある二つの異なる音波が耳に届くとき、弱い音波は強い音波に打ち消されてしまうという聴覚特性のことである。例えば、車などが通り過ぎる時や音楽を大音量でかけている時に誰かに話しかけられても中々聞き取ることができない。この特性は、視覚においても同様で、画素値同士の差が小さい画像に変化を加えることに比べ、画素値同士の差が大きい画像に変化を加えた場合の方が、画素値の変化が周囲に紛れ、画像の変化が目立ちづらくなる。^[39]画像を評価する上で考慮すべき視覚特性として、

1. 空間周波数方向の視覚感度（高周波成分に加わる雑音ほど検知されにくい）
2. 雑音マスキング効果（変化の激しいブロックでの雑音は検知されにくい）が挙げられる。これらは、画像に加わる雑音の程度が、元画像の変化の度合いによって異なることを示唆している。1のように、変化の激しい領域に重畳する雑音は検知されにくい。2では、低周波成分に加わる雑音でも、その周辺ブロックが高い周波数成分を含むほど、雑音を強く覆い検知されにくい事を

^[33]
意味する。

図 26 にマスクング効果小、図 27 にはマスクング効果大のイメージ図を示す。

5.2.5 PSNR (ピーク信号対雑音比)

原画像となにかしらの処理をした画像を比較したときのノイズの割合のことを PSNR (ピーク信号対雑音比) ^[40] という。PSNR40 [dB] 以上で元の画像と見分けがつかない、PSNR35 [dB] 以上で透かし入り画像としては良しとされるレベルであるとされている ^[33]。本研究では透かしの評価を PSNR で行った。以下に PSNR の計算式を記す。

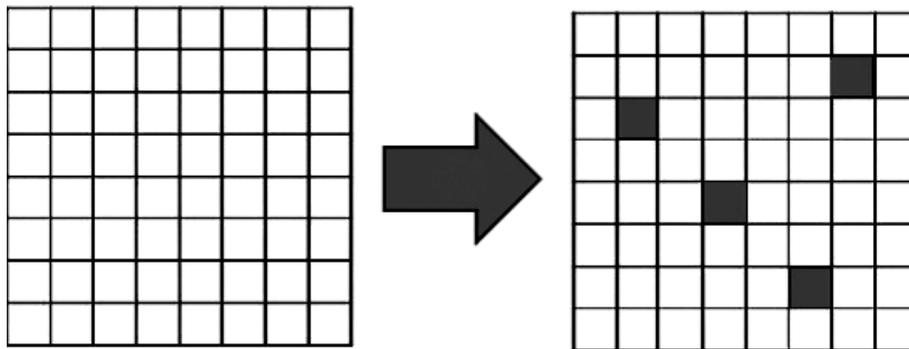


図 26 マスクング効果小

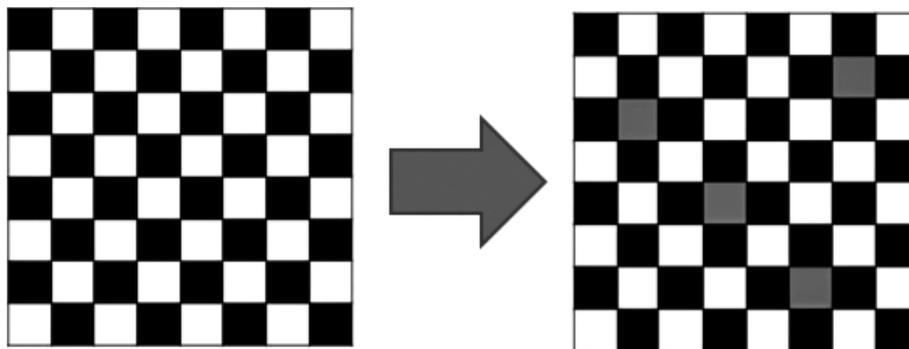


図 27 マスクング効果大

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (x(i, j) - x'(i, j))^2 \quad (5)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (6)$$

単位はデシベル [dB] となる。MAX は原画像がとりうる最大画素値のことである。また、この式の中で表される MSE とは平均二乗誤差 (Mean Squared Error) のことである。m, n は画像の縦、横のサイズを指し、x は原画像、x' は処理画像を指す。

5.2.6 MATLAB

MATLAB は、行列計算、ベクトル演算、グラフ化などを行う数値解析ソフトウェアの一種である。MATLAB を用いることで、C 言語や FORTRAN といったプログラミング言語よりも簡単に計

算を行うことができる。^[45]本研究では MATLAB R2015a を用いて研究を行った。

5.3 提案手法

5.3.1 提案方式

提案方式の流れを以下に示す。

1. ハフ変換
2. DCT
3. 透かし情報の埋め込み
4. IDCT
5. 検出

ハフ変換 元画像にハフ変換を行い、画像の中の最も長い直線を抽出する。本研究において、 $3,685 \times 4,124$ の画像にハフ変換を行った。その結果を図 28 に示す。

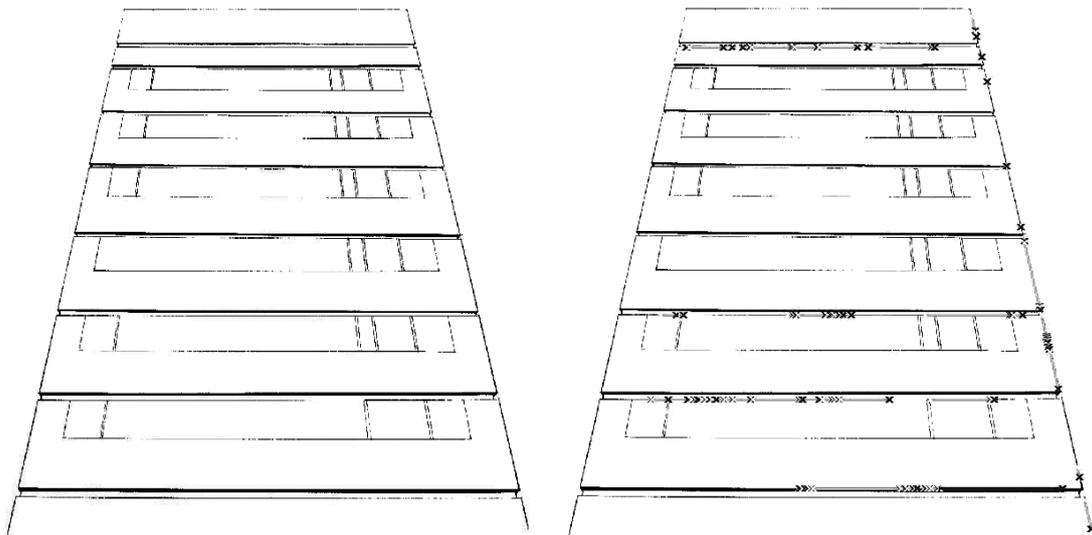


図 28 元画像と埋め込み後の画像

図 28 の左側が元画像、右側がハフ変換後の画像になっている。右側について、黄色の x が始点、赤色の x が終点となっており、その間の直線を緑色で表している。また、最も長い直線を赤い色の線で表示している。

DCT ハフ変換で求められた始点から終点までの検出された直線に DCT を行う。この時、画像のサイズ、透かし情報の桁数によってブロックの数を変化させる。本研究では $3,685 \times 4,124$ の画像の中の抽出された最も長い直線 1×831 に埋め込む。手法として直線を 5×5 のブロックに変換して、一つのブロックに対して白と黒の画素値を埋め込むこととする。この時、DCT 係数の値を X 、埋め込み後の係数の値を X' 、埋め込む透かし情報の値を Y 、埋め込み強度を $A (> 0)$ とすると、埋め込む時の場合分けは以下の式 (3.1) になる。

$$X' = \begin{cases} 1 \times A & (Y = 1) \\ 0 & (Y = 0) \end{cases} \quad (7)$$

ハフ変換された直線を5×5のブロックに変換する方法については以下に示す。1×5の5ブロックずつに分割する。

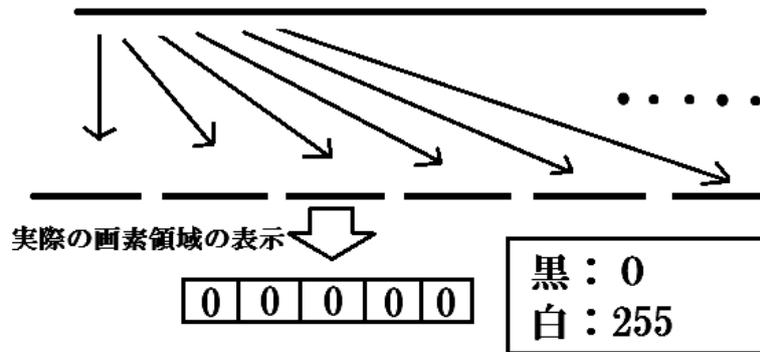


図 29 イメージ図

分割された5ブロックの線を5本縦に並べて5×5のブロックを作成する。そして、作成された5×5のブロックにDCTを行う。

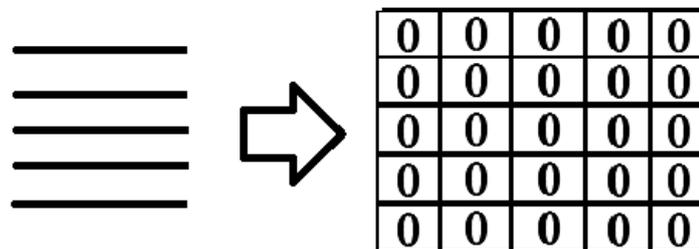


図 30 イメージ図

透かし情報の埋め込み 透かし情報には二進数の数字を利用する。本研究では、学生番号 201202750 の十進数を二進数に変換して0と1で構成された28桁を透かし情報に用いた。埋め込む際にはブロックに対して、1と0のどちらかの値を埋め込むことにする。また、埋め込む際に埋め込み強度を指定する。埋め込み強度が大きいほど埋め込み強度が高いことを意味する。

IDCT 透かし情報が埋め込まれた5×5のブロックを周波数領域から画素領域にIDCTで変換する。その後、直線に戻して透かし入りの画像の完成となる。

検出 検出はDCTを行い、透かし情報の埋め込まれた画像から検出する。方法は埋め込みの手順と同様に行い、透かし情報を得る。Yを検出結果とし、DCTで埋め込まれた透かし情報の取り出しは式(3.2)になる。

$$Y' = \begin{cases} 1 & X' \div A \\ 0 & X' \div A \end{cases} \quad (8)$$

5.3.2 結果

今回使用した元画像および埋め込み後の全体画像を図 31 に示す。

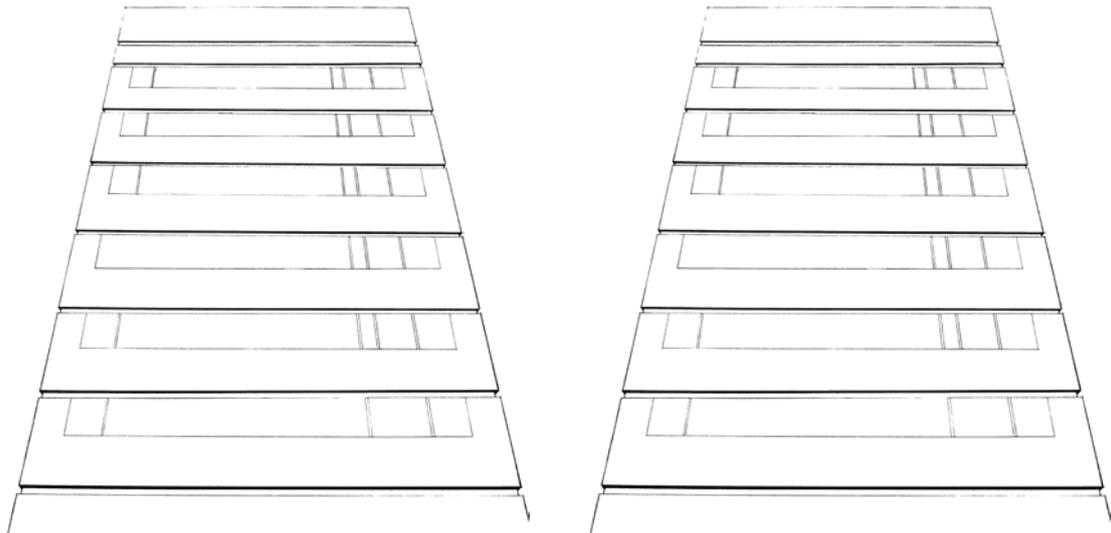


図 31 元画像と埋め込み後の画像

図 31 は元画像で、右図は埋め込み後の画像である。図 31 には今回 28 桁の二進数の情報が入っている。

次にハフ変換で抽出された画像の埋め込み前と埋め込み後の画像を図 33 に示す。

埋め込み前 抽出画像

図 32 ハフ変換の埋め込み前画像

埋め込み後 抽出画像

図 33 ハフ変換の埋め込み後画像

今回は研究中に埋め込み情報が精度良く検出できなかったので、低周波に近い部分での埋め込みを行った。次に透かし情報の検出結果を以下に示す。

検出結果

二進数の桁数	1	2	3	4	5	6	7	8	9	10	11	12	13	14
埋め込み情報	1	0	1	1	1	1	1	1	1	1	1	0	0	0
検出(少数 4 桁)	0.6699	-0.1325	0.4998	0.4991	0.5993	0.5007	0.4998	0.5007	0.4998	0.4998	0.5007	0.1317	0.0000	-0.0009
二進数の桁数	15	16	17	18	19	20	21	22	23	24	25	26	27	28
埋め込み情報	0	1	1	1	0	0	0	0	1	1	1	1	1	0
検出(少数 4 桁)	0.0000	0.4998	0.4998	0.4998	0.0000	0.0000	0.0000	-0.0009	0.4989	0.6394	0.5996	0.7173	0.4824	0.0000

5.3.3 透かし情報の劣化についての考察

検出された情報は0に近似された値と0.4以上の値の二種類となっている。0.4以上の値を1とするよう条件付ければ埋め込んだ情報と一致する。しかし、検出時では透かし情報は劣化している。原因としてはDCTを使用する埋め込み位置と埋め込み方法に原因があると思われる。また、今回埋め込み強度係数を256で行ったが、最適な値ではない可能性がある。

5.3.4 元画像と比較する考察

元画像と埋め込み画像でPSNRを使って算出した。結果としては45.1726という値を示した。理論上、抽出された直線以外は変わっていないため、画像全体で見た場合は変化が少なく見える。しかし、抽出された部分だけで見れば変化が分かるため、埋め込み場所に気付かず注視しなければ肉眼では分かりづらいと思われる。

次に今回容量での比較を検討した。理由として元画像の容量は1,183,511バイトであり、埋め込み画像の容量は566,433バイトであったためである。

分類	元画像	埋め込み画像	埋め込み画像
画像のサイズ	3,685×4,124	3,685×4,124	3,685×4,124
容量(バイト)	1,183,511	566,393	566,433
およそ	1.18MB	0.57MB	0.57MB

本来、埋め込み画像は透かし情報を含むので元画像より容量が増加するのが一般的である。しかし、埋め込み画像の容量が元画像の約半分となっている。つまり、半分近くの情報が破棄されている。抽出された直線以外は変化しておらず、直線の変化だけでこの容量差になるとは思われない。考えられる要因としてMATLABの処理で画像をデータで読み込み、そのデータを画像として保存した一連の動作による影響だと考えられる。

そこで元画像を読み込んだ場合での比較も行った。読み込み画像と埋め込み画像では容量が近似しており、埋め込み画像の方が微かに大きい。これは透かし情報による差だと思われ、前述した条件にも合致する。このため、読み込んだ画像を参照イメージとしてPSNRをした時の値も変化する。場合分けをしたPSNR比較の結果を以下に示す。

PSNR 参照イメージ	元画像	読み込み画像
値[dB]	45.2	61.1

5.4 結論と今後の課題

本研究でハフ変換とDCTを用いて直線の周波数成分に埋め込む電子透かしの手法を提案した。目的は画像の劣化を抑える電子透かしである。結論として、今回の提案手法は埋め込みに成功しているが検出された情報は劣化してしまった。また、画像の劣化としてはPSNRでの比較を検討した結果、PSNRの値は元画像および読み込み画像の両方ともに40[dB]以上の数値を示した。その上で1ドットの直線に埋め込んだため、人間の目では確認しづらい。つまり、画像の劣化を軽減できた。今後の課題としては幾つか挙げられる。

一つ目は汎用性の向上である。今回ハフ変換で抽出された直線は横向きの線であり、それに合わせた透かし情報の埋め込み処理を行った。そのため、縦向きの線が抽出された場合は埋め込み処理を再構築しなければならない。

二つ目は埋め込み方法の改善である。透かし情報が劣化してしまったため、埋め込み位置に値を代入する以外の埋め込み方法を変える必要がある。また、強度係数の最適化するアルゴリズムが必要だと思われる。

三つ目は先行研究である P 型フーリエ記述子との比較を行う。

四つ目は本研究で DCT の二次元で行ったが一次元で行っていないため、どちらが適切か比較が必要だと思われる。

6 利他的効用関数を用いたネットワーク市場取引のモデル化 [目的 3]

6.1 序論

6.1.1 背景

今日、多くのものが法定通貨によって取引されている。また、インターネット等のネットワーク技術の発達により、場所を選ばずに誰とでも容易に個人間の売買が可能となった。誰でも気軽に売買を行えるということからインターネット上での物の売買が増えているが、インターネットという性質上、取引の相手は不透明であり、どんな人物かわからないことが多い。そのため直接取引を行うよりも容易に相手を脅すことができ、金銭や物を脅し取るといった犯罪が増加しており、インターネット売買での秩序の崩壊が一つの問題となっている。ここでいう秩序とは法やルールのことであり、秩序の崩壊している状態とは多くの人が法やルールを守っていない状態のことを指す。

6.1.2 問題提起

これまでに、より良い価値交換を行うために様々な研究がされてきたが、その多くがプレイヤーは合理的であり、自分自身の利得を最大にするために行動をすると仮定されている。研究では多くの場合、金銭を客観的な利得としてそれを最大にするために行動をする。確かにこの行動は自然なものだが、人は常に客観的な利得のために行動をしているとは限らない。実際に現実の世界では、ボランティアや献血、募金など自分以外の誰かのためだけに行動することや、逆にいじめや暴力など自分に直接の利益にならない行動を取ることもある。このことから、より現実に近い価値交換のモデル化を行うためには、金銭や物といった客観利得以外にプレイヤー毎に違うであろう価値観を考慮する必要がある。

6.1.3 研究目的

問題提起で述べたように、現実の世界では金銭だけを利得とし合理的に行動していると仮定すると、説明できないような行動を行っている（ボランティアや募金など）。そこで本研究では、このような行動を利他的行動と呼び、利他的行動は利他性の有無による行動と考える。「利他性」という概念を取り入れることによって、より現実に近い価値交換のモデル化を行う。また、利他性を考慮した

利他的効用関数を用いてコミュニティ全体の利他的な満足度を上げていったとき、コミュニティ内の平均の利他性はどのように変化していくのか、シミュレーションを行い探っていく。利他性とは自分の利益よりも相手の利益を優先して行動することであり、利他性が高いほど相手のために行動をする。

6.1.4 提案

本研究ではコミュニティ内のプレイヤー達の利他性を考慮し、取引ができるような価値交換システムを提案する。想定モデルとして、お互いに面識のないインターネット上での売買で、売買に参加するプレイヤーが入れ替わっていくという開放的なコミュニティを想定する。各プレイヤーにあらかじめランダムに利他性を与え、その利他性の値から「強く利他的」「協力的」「やや利他的」「純粋に利己的」「やや攻撃的」「競争的」「強く攻撃的」の七つに分類する。この分類した利他性を用いて、あらかじめ作成した利得表からプレイヤー対コミュニティで取引を行う。人の利他性は、感情や状況、行動によって変化していくものと考えられるため、本研究では取引の結果を考慮し、遺伝的アルゴリズムを用いて各プレイヤーの利他性を変動させていく。この利他性による分類から、遺伝的アルゴリズムを用いた利他性の変動までの流れを1世代とし、世代を繰り返すことで、コミュニティの平均の利他性がどのように変化していくのか探っていく。

6.2 基礎知識

6.2.1 価値

用途の異なるもの（商品）を交換するときに、その交換を可能にする基準となるもの、通常は貨幣に換算した価格で表される。商品は使用価値と交換価値を持ち、使用価値は人間にとって有用なその商品の使い道で、交換価値は商品を交換する当事者にとっての価値である。この交換価値を全ての商品で普遍的に成り立たせている物が価値と呼ばれ、アダム・スミス以来の経済学の重要なテーマになってきている。価値論は大きく二つに分かれ、一つはアダム・スミスからリカードを経てマルクスによって完成された労働価値説、もう一つはメンガーやワルラスが提唱した主観価値説^{[56][55]}である。

労働価値説 労働価値説は、商品生産に投入された労働を価値の実体と見なす。マルクスはこの労働を使用価値を生む具体的な有用労働に対して、抽象的人間労働と呼んだ。これは具体的な労働の内容を考慮せず、誰もが普通にできる労働を基準にしているので、複雑労働を単純労働に換算できる。価値の大きさは労働の量、つまり労働時間で計られる。

主観価値説 主観価値説は、商品を買って使用することから得られる満足を価値の実体と考える。あらゆるものの価値は、市場でそれがどれだけ消費者の欲望を引き付けるか、また希少性が高いか、の2点で決定される。

6.2.2 秩序問題

秩序問題とは、「自由な個人と社会秩序はいかにして両立するのか」という問題である。「自由な個人」と「秩序」のについての考え方はいくつかあるが、ここでは秩序問題を「ホッブズの秩序問題」

と「日常世界的秩序問題」の二つに大別する。^[48]

ホッブズ問題 もし国家や法律がなく、何の拘束もない人間の集合とそのような状態における人間の諸性質として仮定したとき、一定の性質を持った個々の人間同士の相互作用は、全体としてある状態を呈する。その状態を自然状態という。ホッブズはこの自然状態を「万人の万人に対する争い」、つまり戦争状態であると導いた。つまりホッブズ問題とは、自然状態における戦争状態を回避し、いかにして社会秩序を形成するか？という問題である。^[47]「秩序」とは「平和」であるという考え方である。ホッブズ的**社会秩序問題**は、「秩序＝平和」という見方を「秩序＝公共財が供給された状態一般」に拡張されたもので、単純化すると、囚人のジレンマの相互協力状態が公共財が十分に供給された状態である。しかし、社会学的には「公共の利益」の中身が問われなければならない、また、秩序一般を公共財に還元することは難がある。なぜなら、日常会話やスポーツ等の秩序はそれ自体としては公共的ではないからである。^[48]

日常世界的秩序問題 日常世界的秩序問題では、日常世界（commonsense world）自体を秩序と見なす。しかし、日常世界そのものの定義が曖昧であり、日常世界の定義を形式化する必要がある（日常世界自体を秩序とすると、下克上が日常化した戦国時代にも秩序が存在したことになる）。^[48]

6.2.3 利他性

利他性とは、他者の客観的な利得の増大が、主体自身にとっても満足をもたらすような主体の心的性質のことである。つまり、他者にとってのプラス（マイナス）が、自分にとってのプラス（マイナス）であると言い換えることができる。

利他性の考慮 現実の世界では、ボランティアや募金のようにその行動をしたことにより、自分の利益にならない、むしろコストがかかるにもかかわらず、他者のために行動するという事実がある。これは、人間がなんらかの動機により行動を起こすと考えるならば、上記で説明した利他性を考慮しなければこれらの社会現象を説明することができない。^[48] よって、これらの行動を利他的行動といえることができる。利他的行動の定義は次項に示す。

利他的行動 利他的行動とは以下のように定義されている。^[48] 暴力や意地悪といった人を傷つける攻撃行動に対し、人や社会のためになり、社会において価値のあるものと認められる行動は向社会的行動（prosocial behavior）といわれる。その中でも、人助け行動のような苦しい状況にある他者を助けるために、ある程度の出費を覚悟して自発的に行う行動を援助行動（helping behavior）と呼んでいる。また、「愛他的行動（利他的行動ともいわれる）」とは、一つの型の「向社会的行動」である。他者のために自発的に行われる行動の動機が純粹に他者のためだけを思ってなされる場合、その行動を愛他的行動（altruistic behavior）というのである。

効用関数 一般的に利得の代表は金銭である。通常、自身の得る金銭が多いほど満足は大きくなると

考えられるが、これ以外にも食べ物の量や、土地の広さなど数量一般に拡張できる。そこで、主体の満足度のことを一般に効用、連続的な尺度を持つ効用を基数的効用と呼ぶ。基数的効用は、数量の関数として記述できることから、それは効用関数として表現される。本研究では、主体が他者に対して利他的な行動を行った場合に得られる満足度を数値的に表した関数を使用し、それを利他的効用関数と呼ぶ。

6.3 価値交換

6.3.1 価値交換の流れ

本研究では、図 34 のような流れで価値交換を行っていく。

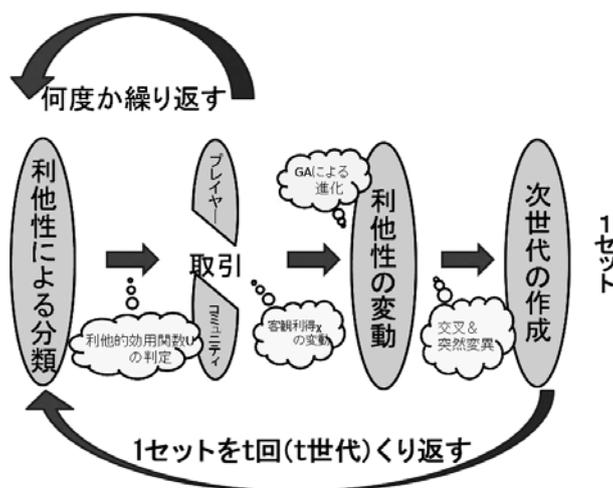


図 34 価値交換の流れ

利他性による分類 本研究では各プレイヤーにランダムに 0~100 の利他性を与え、その利他性の値からプレイヤーを「強く利他的」「協力的」「やや利他的」「純粋に利己的」「やや攻撃的」「競争的」「強く攻撃的」の七つに分類していく。この分類した七つの利他性を用いて取引を行う。

取引 プレイヤー対コミュニティでの取引を行う。コミュニティの利他性は全プレイヤーの平均の値とし、図 35 の利得表を用いて取引を行う。取引は 1 世代のうちに全プレイヤー一度ずつ行う。

利他性の変動 人は誰しも利他性を持っていると考えられる。しかし、利他性には個人差があり、またその時の感情や状況によって人の利他性は変化していきだろ。イライラしているときには普段よりも利他性は低くなり、誰かをいじめたくなるかもしれない。いいことがあった後には自分も誰かにいいことをしてあげようと思うこともあるだろう。そこで本研究では、取引の結果により各プレイヤーの利他性を変動させていく。方法としては、遺伝的アルゴリズムを使用する。遺伝的アルゴリズムについては、節 6.3.3 で詳しく説明する。

6.3.2 利他的効用関数

本研究では各プレイヤーに客観利得 x_i^t と利他性 e_i^t をランダムに与える。ここで利他的効用関数を

コミュニティ プレイヤー	強く利他的	協力的	やや利他的	利己的	やや攻撃的	競争的	強く攻撃的
強く利他的	4,3	2,4	0,5	(-2),4	(-5),3	(-7),3	(-9),3
協力的	5,2	3,3	1,4	(-1),3	(-4),2	(-6),2	(-8),2
やや利他的	6,1	4,2	2,3	0,2	(-3),1	(-5),1	(-7),1
利己的	7,0	5,1	3,2	1,1	(-2),0	(-4),0	(-6),0
やや攻撃的	8,(-1)	6,0	4,1	1,0	(-1),(-1)	(-3),(-1)	(-5),(-1)
競争的	9,(-2)	7,(-1)	5,0	2,(-1)	0,(-2)	(-2),(-2)	(-4),(-2)
強く攻撃的	10,(-3)	8,(-2)	6,(-1)	3,(-2)	1,(-3)	(-1),(-3)	(-3),(-3)

図 35 利得表

U_i^t とし、以下のように定義する。

$$U_i^t = v(\theta_i^t)\chi_i^t + \{1 - v(\theta_i^t)\}\chi_i^t$$

$v(\theta_i^t)$ は自己の客観利得の重みである。つまり、利他的効用関数とは自分が得た利得による自分の満足度と、相手が得た利得による自分の満足度を足したものである。また、コミュニティ内の平均の利他的効用関数を U_a 、コミュニティ内の平均の利他性を e_a とする。

6.3.3 遺伝的アルゴリズム

遺伝的アルゴリズムとは、生物が環境に適応して進化していく過程を工学的に模倣した学習的アルゴリズムである。自然界における生物の進化過程では、ある世代を形成している個体の集合の中で環境に適応した個体が高い確率で生き残り、次の世代に子を残す。このメカニズムをモデル化し、環境に対して最もよく適応した個体、つまり目的関数に対して最適値を与えるような解を求めようというのが遺伝的アルゴリズム (GA) の概念である。

遺伝的アルゴリズムの流れ 一般的な遺伝的アルゴリズムのフローチャートを以下に示す。^[54]

- 初期集団の生成 あらかじめ設定された数だけ個体を生成する。生成した個体の数のことを母集団サイズ (Population Size) や単に個体数 (Number of Individuals) と呼ぶ。
- 評価 (Evaluation) 各個体の持つ染色体を問題空間にデコードして解を求める。ここで求めた解

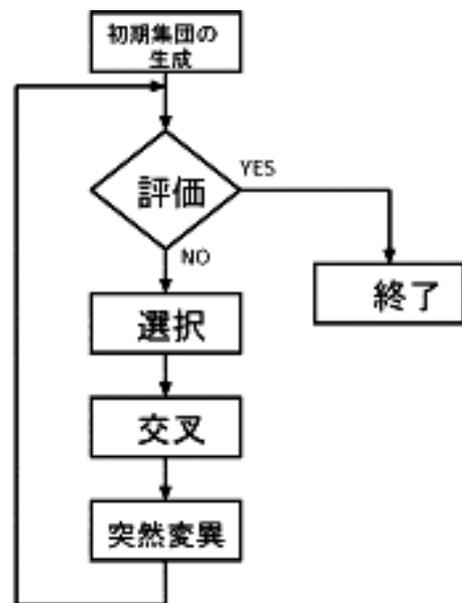


図 36 遺伝的アルゴリズムのフローチャート

のことを評価値 (Evaluation Value) という。

- 選択 (Selection) 生物の適者生存を模倣したものである。この操作では、まず各個体の評価値から次世代への生き残りやすさを求め、これに基づいて次世代の母集団を形成する。この次世代への生き残りやすさのことを適合度 (Fitness) と呼ぶ。
- 交叉 (Crossover) 生物の有性生殖を模倣したものである。この操作により、個体間で染色体情報が交換される。最適解を表す個体の一部分を持った個体同士が交叉すればより最適解に近い個体が得られる可能性が高くなる。個体集団のうち何割の個体が交叉するかを交叉率 (Crossover Rate) と呼ばれるパラメータによって定める。本研究では、交差点を1箇所選んで、その前と後で遺伝子を入れ替える「1点交叉」という方法を用いる。
- 突然変異 (Mutation) 染色体は遺伝子 (Gene) を格納する複数の遺伝子座 (locus) から構成され、遺伝子座に入りうる遺伝子のことを対立遺伝子という。突然変異とは、染色体上の遺伝子座の遺伝子を別の対立遺伝子に置き換える操作のことであり、自然界における DNA 複写の際に起こるコピーミスにあたる。染色体のうち何割の遺伝子座が変異するかを突然変異率 (Mutation Rate) と呼ばれるパラメータによって定める。
- 終了判定 (Terminate Check)
あらかじめ定められた終了条件に基づいて遺伝的アルゴリズムを終了させる。

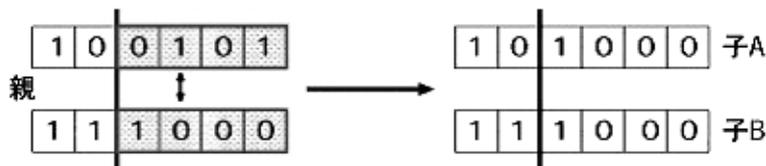


図 37 1点交叉の例

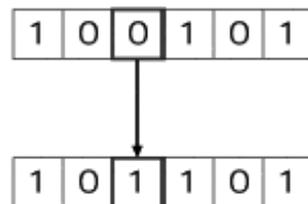


図 38 突然変異の例

遺伝的アルゴリズムの特徴 Goldbergによると、遺伝的アルゴリズムは従来の最適化手法と比較して四つの特徴があるといわれている。

- 設計変数を直接操作せずにコード化した状態で扱う
- 一点探索ではなく、多点探索である
- サンプリングによる探索で、ブラインドサーチである
- 決定論的規則ではなく、確率的オペレータを用いる探索である

6.4 シミュレーション

今回の研究では以下の状況を考え、シミュレーション実験を行った。

- プレイヤーが入れ替わり、新たなプレイヤーが取引に参加する開放的なコミュニティ
実際にシミュレーションを行った結果をいくつか以下に示す。

例1：利他性をランダムに与えた場合（新たに参加するプレイヤーの利他性もランダム）

表3 初期設定値

プレイヤー数	100
客観利得 χ_i^t	0~100
利他性 θ_i^t	0~100
世代数 t	5000

表4 例1のシミュレーション結果

U_a	2.62×10^5
θ_a	45.8
コミュニティの分類	純粋に利己的

表5 例2のシミュレーション結果

U_a	2.61×10^5
θ_a	46.3
コミュニティの分類	純粋に利己的

表4のシミュレーション結果を見ると、ばらつきながらも多くの世代でコミュニティ内の平均の利他性は40~50の間に集まっていることが分かる。表5からコミュニティの分類は「純粋に利己的」である。つまり多くのプレイヤーが、必要以上に相手の利得を奪おうとしない程度で、自身の利得を最大にするために行動をしているといえる。

例2：コミュニティ内の9割のプレイヤーが強く利他的（新たに参加するプレイヤーの利他性はランダム）

図40では、数回の世代交代を繰り返したのち、図39同様、コミュニティの平均の利他性は表5から純粋に利他的となった。この結果は、多くの利他的なプレイヤーが少数の攻撃的なプレイヤーに利得を奪われた結果、利他的なプレイヤーがコミュニティを離れたためと考えられる。

例3：コミュニティ内の9割のプレイヤーが強く攻撃的（新たに参加するプレイヤーの利他性はランダム）

図41では、数回の世代交代を繰り返したのち、図39同様、コミュニティの平均の利他性は表6から純粋に利他的となった。この結果は、多くの攻撃的なプレイヤー同士が取引をした結果、お互いに損をしコミュニティを離れたためと考えられる。

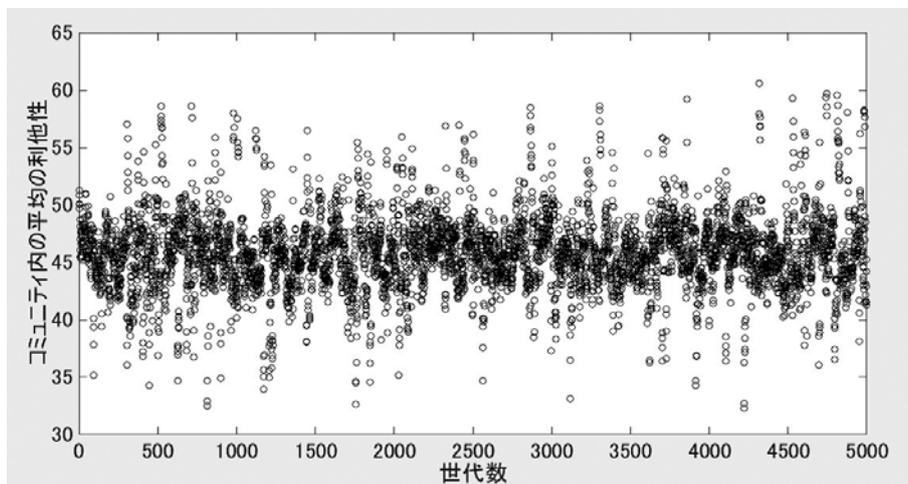


図39 例1のシミュレーション結果

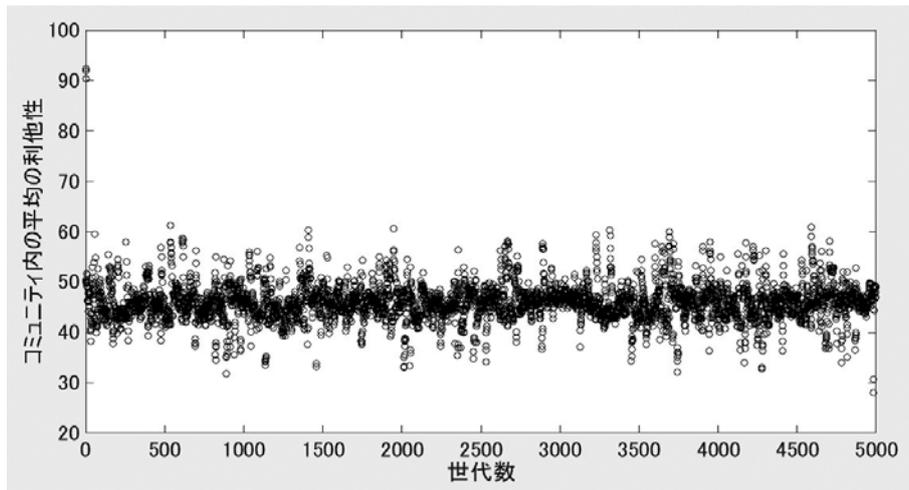


図40 例2のシミュレーション結果

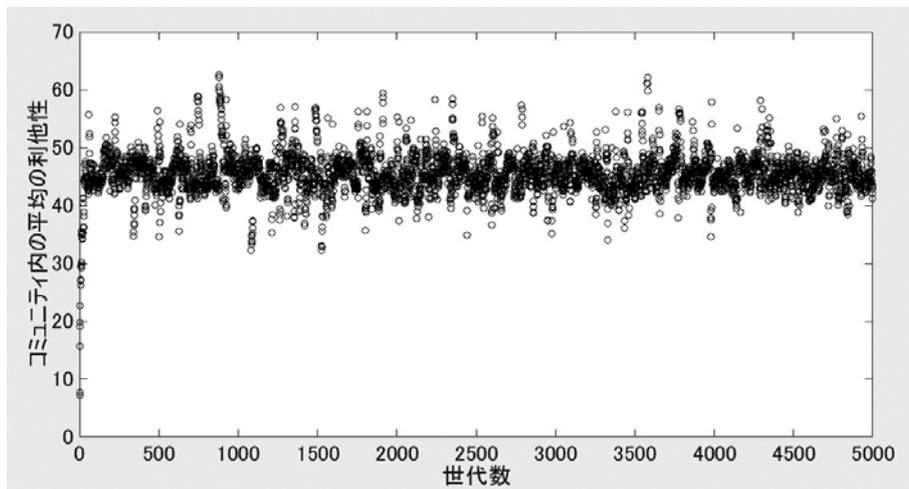


図41 例3のシミュレーション結果

表6 例3のシミュレーション結果

U_a	2.75×10^5
θ_a	45.6
コミュニティの分類	純粋に利己的

例4：コミュニティ内の9割のプレイヤーが強く利他的（新たに参加するプレイヤーの利他性も強く利他的）

図42では、追加プレイヤーを強く利他的としたため、利他性の高いコミュニティを作成することができた。しかし、図39の結果では U_a は 2.62×10^5 だったのに対し、図42の結果では U_a

は 1.52×10^5 となり、利他性は高いにもかかわらず、利他的な満足度は低いという結果となった。

例5：コミュニティ内の9割のプレイヤーが強く攻撃的（新たに参加するプレイヤーの利他性も強く攻撃的）

図43では、追加プレイヤーを強く攻撃的としたため、利他性の低いコミュニティを作成された。表8から、利他性があまりに低くなり、攻撃的なプレイヤーが増えると U_a はマイナスになることが分かる。

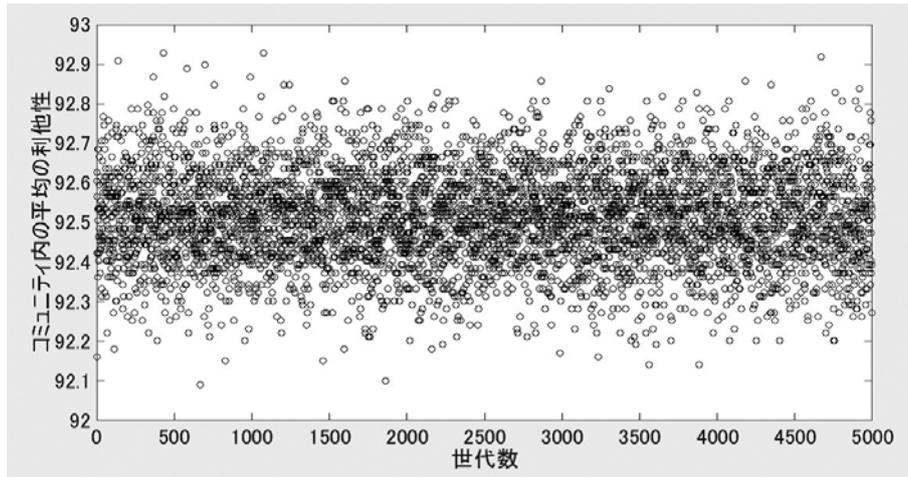


図42 例4のシミュレーション結果

表7 例4のシミュレーション結果

U_a	1.52×10^5
θ_a	92.5
コミュニティの分類	強く利他的

6.5 結論

図39のシミュレーション結果を見ると、ばらつきながらも多くの世代でコミュニティ内の平均の利他性は40～50の間に集まっていることが分かる。表5からコミュニティの分類は「純粋に利己的」である。現実の世界では、ある程度の利他性はあるものの、自分の方が大事であるという「やや利他的」な人が多く存在すると考えられる。しかし、今回のシミュレーションでは利他性による分類は「純粋に利己的」に分類された。これは、今回のシミュレーションがネットワーク上での取引を想定していたためと考えられる。ネットワーク上での売買では、取引の相手は不透明なことが多く、同じ相手と何度も取引を行う可能性も低い。つまり、その取引以外での関わりは少ないため、実際に会って取引を行うよりも利他性は低くなり、自分の利益を優先すると考えられる。だが、だからといって攻撃的な人間が増えるわけではなく、多くの人が相手を傷つけないことを前提とした最大の利得を得るために行動をするだろう。このことから、コミュニティ内の平均の利他性が純粋に利他的となった今回のシミュレーション結果は妥当であり、現実に近いコミュニティのモデル化を行うことができたといえる。また、図40～図42は1世代目のコミュニティ内のプレイヤーの多くが「強く利他的」又は「強く攻撃的」なときにコミュニティ内の平均の利他性はどのように変化していくかをシミュレートしたものである。図40、図41では、新たに参加するプレイヤーの利他性はランダムに与え、図42では追加プレイヤーの利他性は強く利他的、図43では追加プレイヤーの利他性は強く攻撃的とした。

図40、図41ではいずれも数回の世代交代を繰り返したのち、図41同様、コミュニティの平均の利他性は純粋に利他的となった。図40の結果は、多くの利他的なプレイヤーが少数の攻撃的なプレイヤーに利得を奪われた結果、利他的なプレイヤーがコミュニティを離れたためと考えられる。また、図41の結果は、多くの攻撃的なプレイヤー同士が取引をした結果、お互いに損をしコミュニティを離れたためと考えられる。

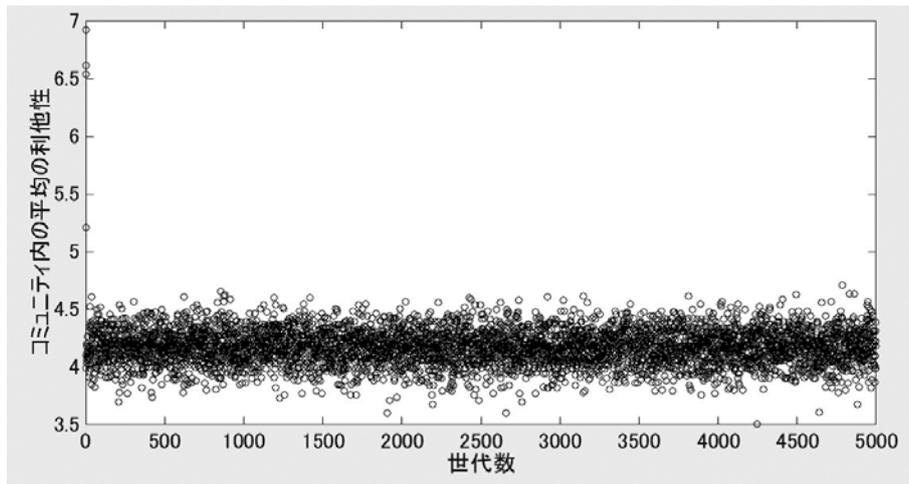


図43 例5のシミュレーション結果

表8 例5のシミュレーション結果

U_a	-1.39×10^6
θ_a	3.9
コミュニティの分類	強く攻撃的

図42では、追加プレイヤーを強く利他的としたため、利他性の高いコミュニティを作成することができた。しかし、図39の結果では U_a は 2.62×10^5 だったのに対し、図42の結果では U_a は 1.52×10^5 となり、利他性は高いにもかかわらず、利他的な満足度は低いという結果となった。これにより利他性が低いほど利他的行動によって

得られる満足度は大きくなると考えられるが、表8から、利他性があまりに低くなり、攻撃的なプレイヤーが増えると U_a はマイナスになることが分かる。つまり、個人の利得を最大にするためには、利己的なプレイヤーが集まるコミュニティよりも利他的なプレイヤーが集まるコミュニティの方が望ましく、個人の利他的行動による満足度を最大にするには、利他的なプレイヤーが集まるコミュニティよりも利己的なプレイヤーが集まるコミュニティの方が望ましいということがわかった。

6.6 今後の課題

今回の研究では、プレイヤーの利得を金銭や物、時間（コスト）というように客観的に確認できるものと定義をしていた。しかし、満足度といった主観的な利得があると仮定し、客観的な利得と主観的な利得を考慮したとき、今回の結果とはまた違った結果が出るものと考えられる。

また、今回はプレイヤーが入れ替わっていく開放的なコミュニティでのシミュレーションを行ったが、同じプレイヤー同士が何度も取引を繰り返す閉鎖的なコミュニティでのシミュレーションではどのような結果になるのか確かめる必要がある。その際、プレイヤー毎に親密度などを設定することで、より現実に近い価値交換のモデル化を行うことができると考えられる。

7 群知能とトピックモデルを用いたファイルシステム [目的4]

7.1 序論

7.1.1 研究背景

近年、スマートフォン、タブレット端末、PCなど個人の持つ電子端末の量が増えている。それに

準じ、一人の人間が管理する情報やファイルの数も増してきている。ファイルの数が増えると管理の手間が増すため、これらの情報を扱うには困難を要する。さらにクラウドやビッグデータといった、大量のファイルを扱うシステムも多く開発され、管理は複雑さを増す一方である。個人や企業が持つPCやスマートフォンなどの電子端末があり、これらにはファイルが入っている。さらに近年では、ネットワークの普及から外部サーバへファイルを預け、またその外部サーバのサービスを基に個人や企業のデータを管理していることも少なくない。

このように情報やデータが端末で分散しているにもかかわらず、これらの情報を管理するためには、各端末で木構造のファイルを管理する必要があるため、これらのデータの場所を把握することは^{[65][66]}難しい。これは一つのファイルが、フォルダという一つの属性しか持たないために引き起こされる、^[67]いわゆる“こうもり問題”が存在しているためである。“こうもり問題”とはこうもりが獣か鳥に属するのかわからず、正しく分類できないという問題であり、これは様々な属性を持ち合わせる今日のファイルのフォルダ分け問題と等しい。ユーザの自分のポリシーやルール（以後「直観」）によって配置したファイルは後に場所を探す場合、直観的に探す場所が分かりにくく、ファイルシステムによって検索をするにも時間がかかるため、ユーザ個人や組織の活動を少なからず阻害している。

これらの多様化するファイルや情報を管理するためのシステムとして、多次元空間で管理を行うものが提案されている。^[68]多次元管理とは、前に述べたように一つの方面でしか見なかったファイル管理の問題を、様々な観点・方向（多次元）から見えるような、管理システムのことである。しかし、この多次元空間での管理では、管理する要素が多いため、ユーザが視覚的に把握するのは難しい。現在、情報漏えい、ファイル改ざんなどの問題があり、これらは他者が重要なファイルにアクセスすることで引き起こされる。人間が決めたものだけでなく、客観的に重要なファイルを見つけ出し、それらを守ることが必要となっている。また、クラウド、ビッグデータといった大量のファイルをプロだけでない多くの人間が管理するためには、より分かりやすいファイルシステムを作ることが、重要であり、情報漏えいの解決にもつながる。したがって、ユーザが情報をより扱いやすくするためには、よりユーザが体感的に扱いやすいファイルシステム、およびそのUI（User Interface）が必要である。

最近では、情報リソースの関係性を表すためのデータマイニングやUIなど視覚的な面の研究も盛んに行われている。^{[69][70][71][72][73]}MicrosoftがWindows Vistaへ搭載するために、WinFSと呼ばれる統合ファイルシステムの開発を行った。また、セマンティックファイルシステム^[74]と呼ばれるタグの概念を用いたファイルシステムの研究も行われている。セマンティックファイルシステムでは複数の要素を一括で管理できるタグの概念を取り入れており、今日でも様々なインターネットサービス内で使われている。しかし、セマンティックファイルシステムは関連性のファイルを探し出すことは容易になるが、目に見える形で整理はされておらず、全体を把握することや、瞬時に目的のファイルを見つけ出すのは困難である。

次に先行研究として二つの研究を挙げる。

①群知能によって、多次元管理を用いることで、作成日、アクセス日時、任意の重要度など様々な属性を反映させたファイル管理を行うことを可能とし、属性ごとの類似度に応じてファイルが集まり、効率よく管理可能とする多次元の手法を提案していた。関連性の強いファイル同士が集まるため、各ユーザにチューニング（その設定や構成を調整し、最大限の性能が発揮できるようにするこ

と) されたファイルの管理を行うことが可能である。^[75]

②群知能の Boid の特徴を用いることでファイルを分類し、重要度の高いファイルが中心に集まる群れを作る研究を行っていた。^[76] 上記の研究ではどちらも、ファイルの使用頻度やそれに伴う他の情報によってファイルの重要性を決めるようにしていた。だがそれだとユーザ自身が必要なファイルであっても使っていないファイルは重要度が低いとされ集まっては来ない。

本研究では上記の様に群知能を使用し、ユーザが本当に必要なファイルを集められるようなシステムを提案する。それにしたがって、従来のファイルシステムにはない「ユーザとシステム間の対話」を取り入れる。

7.1.2 問題点

現在、一般的に用いられているファイルシステムには、ディレクトリが採用されている。これはファイルの入れ物のような構造をしており、グループごとの整理に適している。今日では、ファイルの数が増大し、そのファイルが複数のファイル分けにおける属性を持つようになってきているため管理が困難になってきている。また、自己管理している場合は自分の直観によってファイルを整理している。システム内の検索で探す場合はデータ量が多いため時間がかかる。自分で探そうとする場合、整理した時の直観とは違うことを考えていた場合、手作業でも時間がかかってしまう。さらにユーザが必要なファイルを集められたとしても、関連している内容のものは集めにくい。これは個人や組織の活動を少なからず妨害している。

7.1.3 目的

本研究では従来システムにはない、直観的であり客観的であるファイルシステムの提案を行う。それを実現するためには PC 上の多数のウィンドを使用する。つまり従来のファイルシステムと合わせて使用するファイルシステムである。なぜなら、人それぞれ必要なファイルや情報は変わっており、一つのシステムだけに任せられるものではない。その様々なウィンドとは既存のファイルマネージャに加え、ユーザの仕事場、先行研究でも使われていた群知能、そしてファイル同士のつながりを確立的に表現できるトピックモデルの四つのウィンド全てを使用して必要ファイルを集める。それはユーザ自身の仕事のログ (以後「振る舞い」) と PC による「対話」である。そのため、ユーザの振る舞いのパターンを新しい言語システムとして表現するシステムの提案であり、ファイルマネージャによる木構造の整理、局所相対的類似の群れの構造を相違作用として実装されるものであり、局所相対的故に通用の原理は使用せず Java による実証をするしかない。

7.2 基礎知識

7.2.1 マルチエージェントシミュレータ

マルチエージェントシミュレータとはマルチエージェントシステムの処理を行うアプリケーションのことである。マルチエージェントシステムとは、複数のエージェントから構成されるシステムであり、個々のエージェントやモノリシックなシステムでは困難な課題をシステム全体として達成する。それぞれ異なった判定アルゴリズムなどの特徴 (キャラクタースティック) を持ったエージェントモ

デルを用い（よってマルチエージェントと呼ばれる）、ある一定以上のエージェントを多数設定し、人工社会を構成しそれぞれ特徴の異なるエージェントの相互作用をシミュレーションするシステム。

同じ特徴を持った複数のノードの相互作用によるシミュレーションとは異なり、例えば様々な個性を持った人間社会など一見予測不可能な事象をモデル化し、できるだけ実際に起こりうる状況を再現しようとすることを目的としたシステム。

7.2.2 群知能

群知能とは、アリや鳥や魚のように自然界に存在する生物の振る舞いに着目し、個々の仕事を分散化し自己組織し集合システムを作り出すことにより、最適化問題などを解決するために作られた人工知能の一つである。^[67]

代表的な群知能を以下に記し、定式化を行う。

Boid Boid とは 1987 年にアメリカのアニメーション・プログラマであるクレイグ・レイノルズによって考案・作製された人工生命シミュレーションプログラムである。Boid というモデル名は、鳥もどきという意味の言葉である “bird-android (バード・アンドロイド)” が短くなったことに由来している。<http://members.jcom.home.ne.jp/ibot/boid.html>。Boid の群を実現させる振る舞いは、三つの要素からなり、「衝突の回避」、「速度を合わせる」、「群の中心に向かう」といった三つのルールを規定するだけで鳥の群をシミュレーションすることができる（図 44）。

- Separation（衝突の回避）：近くにいる仲間と衝突しないようにする
 - Alignment（速度を合わせる）：近くの仲間と速度を一致させようとする
 - Cohesion（群の中心（重心）に向かう）：近くにいる仲間を囲まれた状態になろうとする
- 群の一連の動きは、自分自身と仲間との間の距離を最適に保とうとするルールを重要視している。

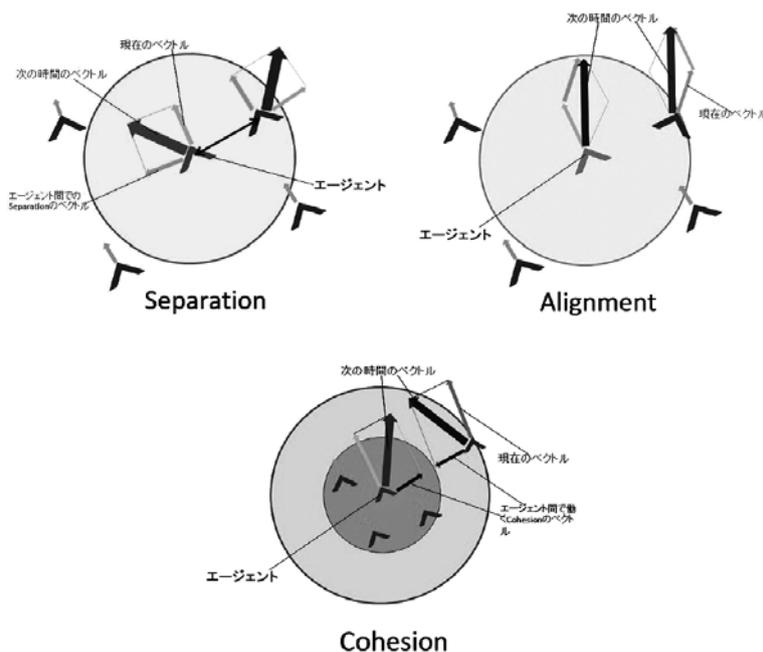


図 44 boid

このような三つの単純な行動規範をそれぞれの個体が持ち、全体として複雑な群の行動が創発する。衝突回避のために、boid はそれぞれ自分にとっての「最適距離」を持っている。自分の最も近くにいる仲間との間で、この距離を保とうと振る舞う。また、速度を合わせるために、最も近くにいる仲間と平行に（同じベクトルで）移動する。これによる速度の変化はない。さらに、群の中心（boid 全体の集合の重心）に向かうようにも速度を常に変更している。

Boid をアプリケーションとして実装している「Mason」を例として挙げる。マルチエージェントシミュレータであり、群知能の動きなどを確認できる。java ベースのアプリケーションである。^[81]

図 45 は Mason 中にある「Flockers」というアプリケーションの起動画面であり、固定数個の個体に対して他の多数の個体が類似度によって集まっている様子を表している。

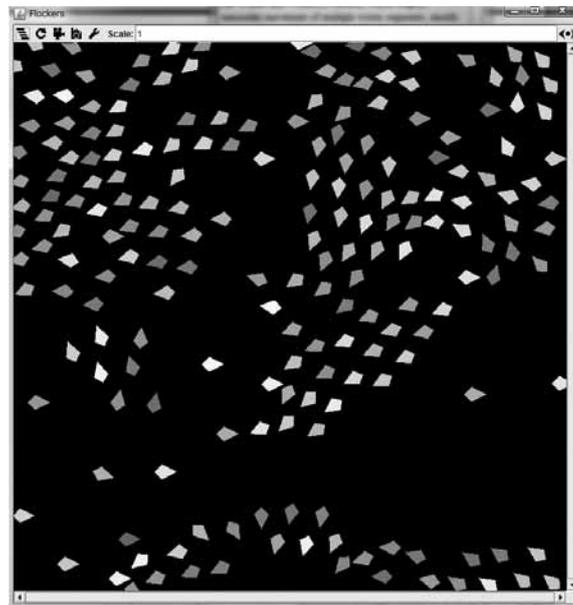


図 45 Mason

ACO (Ant Colony Optimization) アリの摂食行動から着想を得たアルゴリズムであり、フェロモンという揮発性物質を模したパラメータの最適化を行う（図 46）。組合せ最適化やネットワークルーティングなどに応用されている。実世界では、アリは初めランダムにうろつき、食物を見つけるとフェロモンの跡を付けながらコロニーへ戻る。他のアリがその経路を見つけると、アリはランダムな工程を止めてその跡を辿り始め、食物を見つけると経路を補強しながら戻る。しかし、時間とともにフェロモンの痕跡は蒸発し始め、その吸引力がなくなっていく。その経路が長いほどフェロモンは蒸発しやすい。それに対して、経路が短ければ行進にも時間がかからず、フェロモンが蒸発するよりも早く補強されるため、フェロモン濃度は高いまま保たれる。従って、あるアリがコロニーから食料源までの良い（すなわち短い）経路を見つけると、他のアリもその経路を辿る可能性が高くなり、正のフィードバック効果によって結局全てのアリが一つの経路を辿ることになる。群知能を用いて最適解を求めるアルゴリズムの一つである。

PSO (Particle Swarm Optimization) 魚や鳥や昆虫などの生物の検索行動に着目し、James Ken-

nedy が考案したアルゴリズムである。最適化を求める過程での群れる性質を使用し、重要なファイルを中心に集合するのに応用されている（図 47）。複数のパーティクルを座標内のランダムな位置に配置し、ランダムに動く。個々のパーティクルは評価関数を参照し、個々の最適位置を更新する。各パーティクルは新しい適性位置を共有し合い、空間内の新しい位置に遷移する。パーティクル同士の最適位置の更新と共有は何度も行い、群全体としての最適位置を見つける。ACO と同様に群知能を用いて最適解を求めるアルゴリズムの一つである。

7.2.3 トピックモデル

近年、文章や購買履歴などの離散データを解析する手法として、トピックモデルが注目されている。トピックモデルとは、文章が潜在意味（トピック）に基づいて生成される過程を確率的に表現したモデルである。トピックモデルを用いることにより、多様なデータに内在する隠れた構造を抽出できる。^[77]

上記を要約すると、文章が何について記述されているかを推定するモデルである。トピックモデルは適用範囲のトピックを抽出しているだけでなく、文章の補助情報に関連付けさせることや、トピックに相関関係を持たせることができる。これは文章だけでなく、画像、ソーシャルネットワーク、論文の引用関係などに応用ができる。^[78]

これは、単語毎にトピックを選択し1文章が複数の分布から生成されていることを意味する。トピックモデルの式を以下で表す。

$$P(w) = \prod_{k=1}^N \sum_z P(z|\theta)P(w_n|z) \quad (9)$$

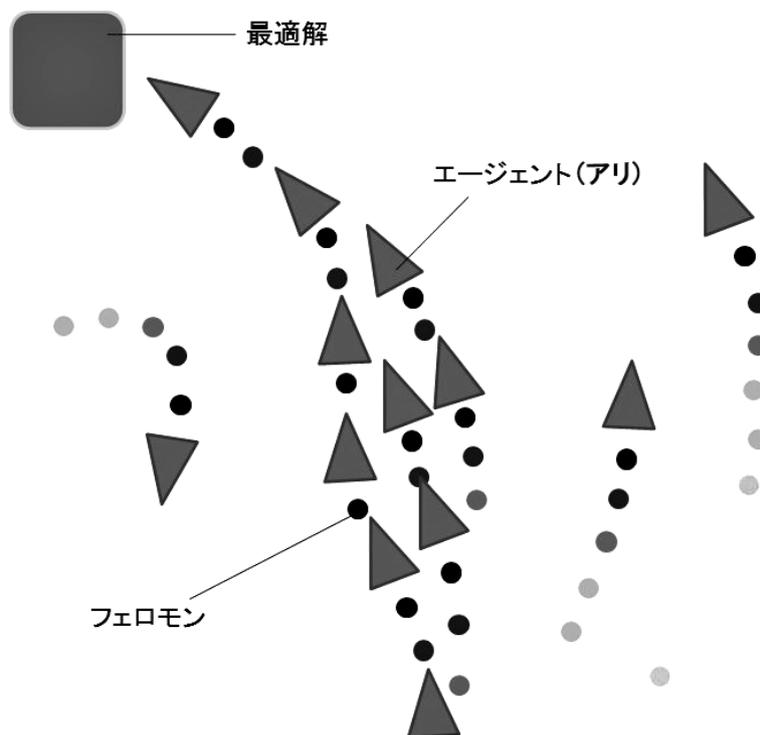


図 46 Ant Colony Optimization

7.3 提案手法

7.3.1 システム概要

ウィンドウとユーザとの対話によって自分にとり最も良い収集を行う。それはユーザとシステムとの対話により自分にとって良い物を探すという意味で最適解を求めるものではない。すなわち、「パターンランゲージ」なのである。

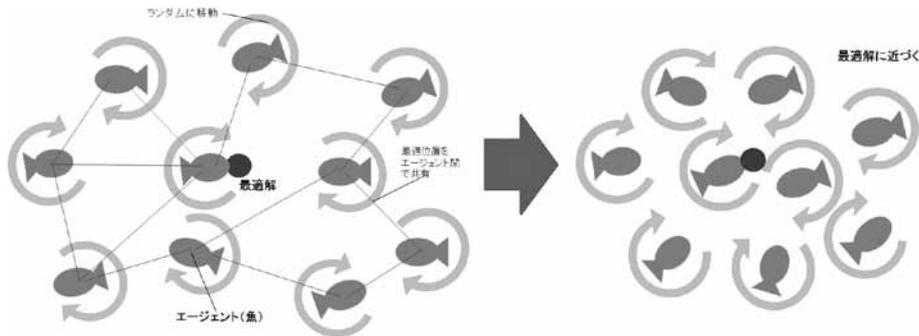


図 47 Particle Swarm Optimization

既存のファイルマネージャ 既存のファイルマネージャによるファイルの収集での問題点を挙げる。例えば、図 48 のように現在ユーザがしている研究の中で、自分の研究用のファイル内にはないシステム A について記されている論文をファイルマネージャ内から探そうとしている。そのユーザはその研究に近いような論文を自分のファイルマネージャの中にランダムに配置している。ファイルの名前だけではユーザの必要な情報であるシステム A について記されているがテキストに入っているか分からない。そのため、ユーザは全文検索機能^[82]などを用いて「キーワード」により、ファイルを集める必要がある。または、ファイルマネージャ内に自分で直観的に配置したものなら、自分の直観に従って探すこともできるが、時間が経つと直観の変化が起こり探すのが困難である。上記の方法などでファイルを集められたとしても、ファイルの中には自分がほしい情報だけで、局所相対的類似でないため情報量は少なく、新たな発見などが生まれる可能性は少ない。

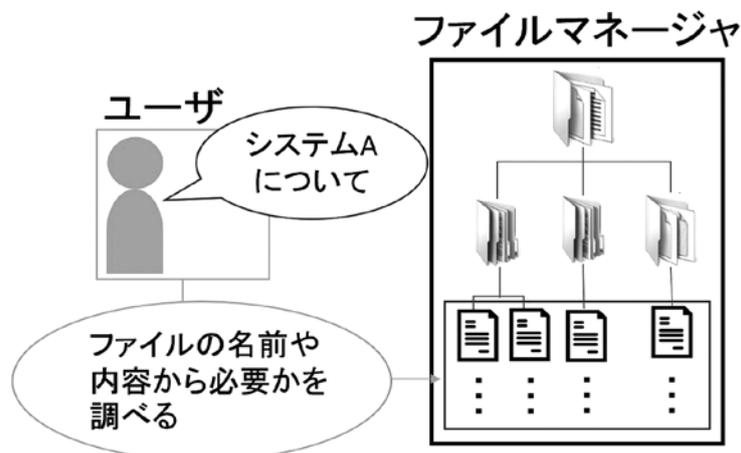


図 48 既存のファイルマネージャ

提案システムにおける各システムの位置づけ 図 49 に提案システムの概要を示す。提案システムはユーザがファイルの振る舞いを把握するために PC と対話するシステムである。

本研究では、Windows OS のディスプレイの中のウィンドウ機能を使用する。①～④は全て別ウィンドウとして表示する。ユーザはそのウィンドウで出た結果を見て各ウィンドウと対話することでより良い結果を抽出する。この対話とはユーザ自身が各ウィンドウに対して自分の意志を反映させることを意味する。

⑤は群知能とトピックモデルを足し合わせたものである。これは本研究の核の部分でもあり、類似度に確率が足し合わされたものでありユーザが思いもよらないファイルが抽出されることが期待できる。

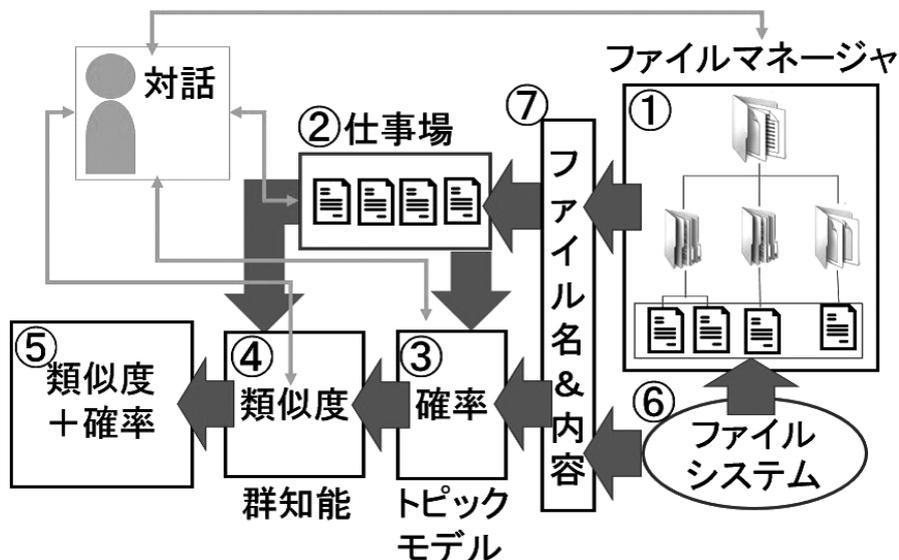


図 49 提案システムの各システムの順路と対話の関係

①ファイルマネージャ

ファイルの移動、削除、改名などができるファイルシステムを扱うためのコンピュータプログラムである。自分が今まで集めた情報がテキストデータとしてファイルの中に、様々な場所に散りばめられて入っている。Mason は仕事場のファイルの類似によって集める。MALLEY はファイルマネージャからトピックを抽出し確率を計算する。

図 50 は上記の一例であり、ユーザのテキストデータの全てとする。このようにユーザは自分の直観の中でファイルをジャンル分けし分散させている。ただしこれは直観により分けてあるだけで場合によっては見方が変わるものもある。

②仕事場（振る舞い）

ファイルマネージャ内にある自分に必要な情報の集合。本研究において自分のしている研究などの作業のことを「仕事」とし、仕事の中で使っている情報の集まりを仕事場とする。

ユーザ自身が現在行っている研究や調べ物などを集めた場所。本研究において仕事場はユーザ自身の意志が最も反映されているところとし、その中に入っているファイルは重要度が高い。ユーザは自分の仕事に関連したファイルをファイルマネージャ内からここに配置する。図 51 は仕事場の例であり、自分に必要なファイルを収納している。

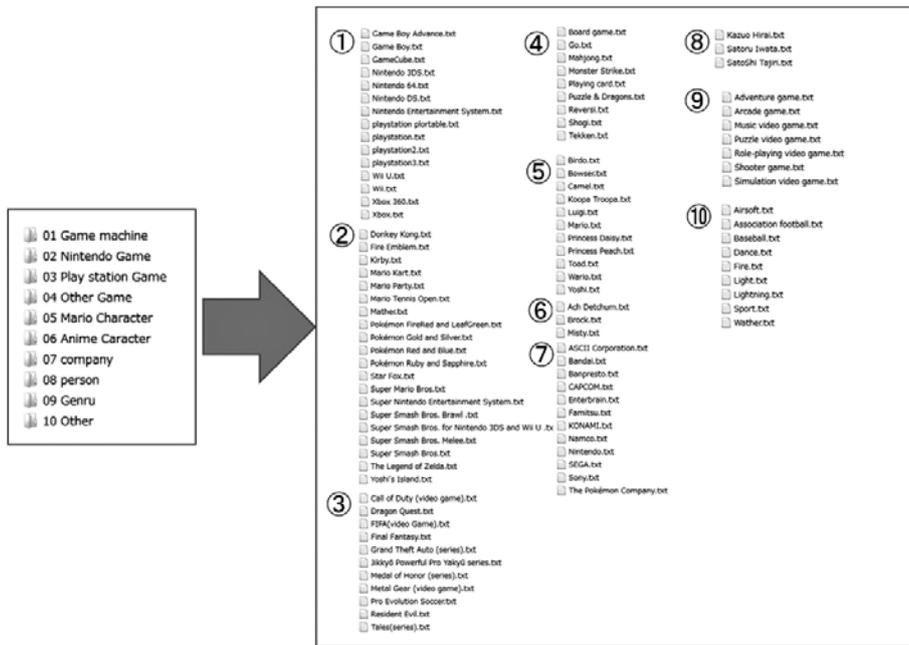


図 50 ファイルマネージャ

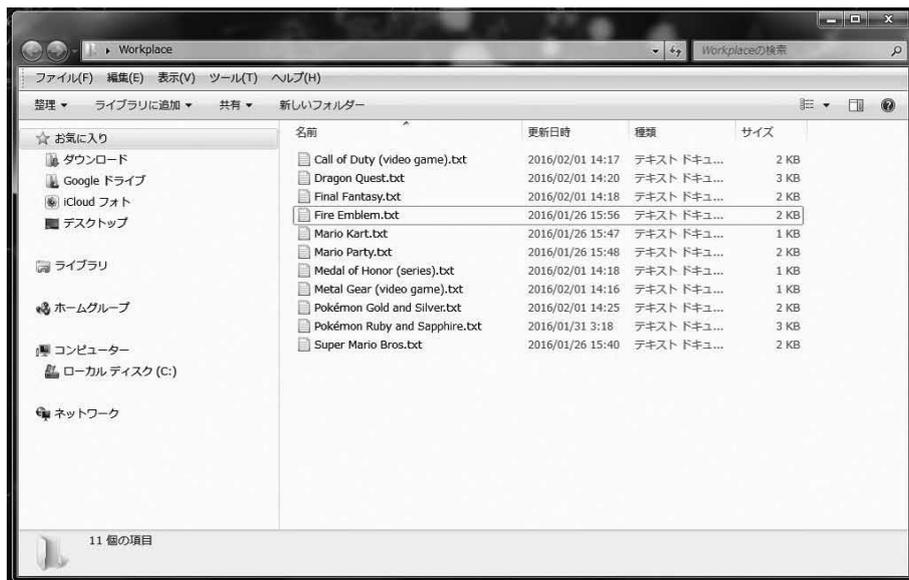


図 51 仕事場

③MALLET

言語処理、文書分類、トピックモデリング、情報抽出、およびテキストに他の機械学習アプリケーション^[80]。大きなテキストの集合体を分析するのに有効で、潜在的ディリクレ分布、階層 LDA などの実装がなされている。Windows 上ではコマンドプロンプトで実行が可能。

群知能の局所相対的類似だけでは自分のファイルマネージャの中から自分の欲しい物全てが見つかるとは限らない。そのため、本研究の核でもあるトピックモデルを動作させるためのアプリケーションとして MALLET を使用する。文章の中に存在する隠れたトピックを抽出する。そのトピックにつ

```

C:\Users\yata>cd desktop
C:\Users\yata\Desktop>mallet import-dir --input text --output data.mallet
Labels =
+ text
C:\Users\yata\Desktop>mallet train-classifier --input data.mallet --trainer MaxEnt --trainer NaiveBayes --train
ing-portion 0.9 --num-trials 100
Training portion = 0.9
Unlabeled training sub-portion = 0.0
Validation portion = 0.0
Testing portion = 0.099999999999999998
----- Trial 0 -----
Trial 0 Training MaxEntTrainer,gaussianPriorVariance=1.0 with 90 instances
Value (labelProb=0.0 prior=0.0) loglikelihood = -0.0
L-BFGS initial gradient is zero; saying converged
L-BFGS initial gradient is zero; saying converged
Trial 0 Training MaxEntTrainer,gaussianPriorVariance=1.0 finished
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 training data accuracy = 1.0
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 Test Data Confusion Matrix
Confusion Matrix, row=true, column=predicted accuracy=1.0 most-frequent-tag baseline=1.0
label 0 |total
0 10 |10
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 test data precision() = 1.0
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 test data recall() = 1.0
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 test data F1() = 1.0
Trial 0 Trainer MaxEntTrainer,gaussianPriorVariance=1.0 test data accuracy = 1.0
Trial 0 Training NaiveBayesTrainer with 90 instances
Trial 0 Training NaiveBayesTrainer finished
Trial 0 Trainer NaiveBayesTrainer training data accuracy = 1.0
Trial 0 Trainer NaiveBayesTrainer Test Data Confusion Matrix
Confusion Matrix, row=true, column=predicted accuracy=1.0 most-frequent-tag baseline=1.0
label 0 |total
0 10 |10
Trial 0 Trainer NaiveBayesTrainer test data precision() = 1.0
Trial 0 Trainer NaiveBayesTrainer test data recall() = 1.0
Trial 0 Trainer NaiveBayesTrainer test data F1() = 1.0
Trial 0 Trainer NaiveBayesTrainer test data accuracy = 1.0
----- Trial 1 -----

```

図52 MALLET

いて確率的にファイルを分析することができる。図52はコマンドプロンプトによる実行画面で確率的リストを表示するアプリケーションである。

④Mason

局所相対的類似を計算するためにMasonを使用する。群知能の似ているものを集めるという役割を持っている。それによってユーザ自身がしている仕事についての情報を仕事場から取得し、それについて類似している物を集める。マルチエージェントシミュレータの群知能による類似のベクトルの定式化したものを以下に示す。言語ゲームにおける“家族類似”を“振る舞い”という位相空間の射の連続性を基に、位相空間の“局所相対的類似”と定義する。

局所相対的類似の振る舞いを表現する群知能ベクトルの公式を以下で表す。類似度 T 、類似度の強度(力) fa 、類似度を強度に変化させる関数 c 、速度 V 、類似度の合計 F 、定数 a 、時間 t とした場合、 i から見た j の類似度の強度、 i から見た類似度の合計、速度ベクトル (i が次にする行動) は

$$fa_{ij} = c_{1(i)}(T_{ij} - T_{(i)t}), T_{ij} > T_{(i)t} \quad (10)$$

$$\vec{F}_{(i)} = a_{1(i)} \cdot \sum_{k=1}^n \vec{fa}_{(i)k} \quad (11)$$

$$\vec{V}_{(i)}^{(t)} = \vec{V}_{(i)}^{(t-1)} + \vec{F}_{(i)}^{(t-1)} \Delta t \quad (12)$$

⑤Mason（類似）+MALLET（確率）の意味

図 49 のようにファイルマネージャからファイル名や内容を用いて、仕事場で使われているファイルとの類似度とそれに対する確率を足し合わせる。

群知能は、〈私〉の仕事場である振る舞いに従って、ファイルの内容やそのファイルの更新情報など類似度によりファイルやデータが群をなして集める機能を持つ。ある属性の観点では、距離の離れた無関係なファイル同士でも視点を変えることにより、着目する属性を変更する前に無関係なファイルと見なされていたものが、類似度の高いファイルと見なされ、群を一箇所に集めることが可能となる。

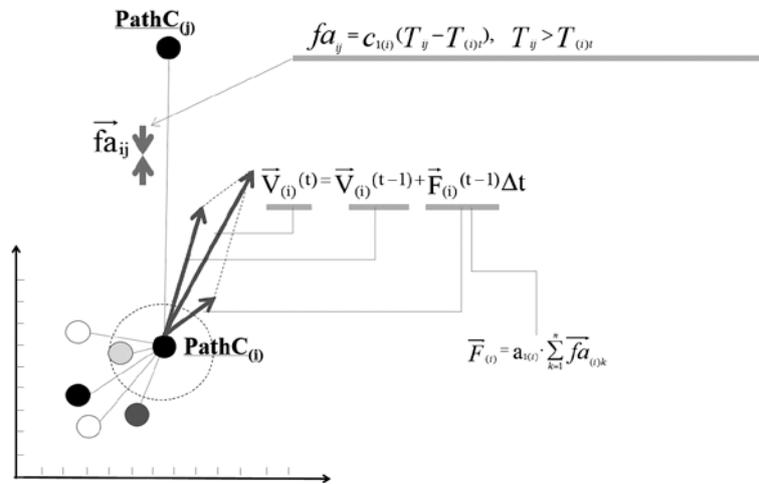


図 53 群知能の定式

- ベイズ：観察結果が得られた場合に、理論（仮説）がどのような確率を持つかを見いだそうとする。
- 頻度：ある理論が正しいとしたときに、観察がどれほど確からしいかを考えるために確率を用いる。^[83]

群知能は、局所的ではあるが理論（類似）を基礎とし観察（テキストを集めること）を正当化する。ベイズ主義は、観察結果が得られた場合に、理論（仮説）がどのような確率を持つか見いだそうとする。ベイズ推定を行うトピックモデルは、そのテキストの集合に潜在的なファイルを引き込むことが期待できる。

式（10）は群知能の公式とトピックモデルの確率を類似強度に変換したベクトルの和を類似度 + 確率とする。式（3.1）第三項はトピックモデルから抽出した引力または斥力であり、その項は、トピックモデルにおけるトピックと、局所相対的なファイルの間の類似を表す。

$$\vec{V}_{(t)} = \vec{V}_{(t-1)} + \vec{F}_{(t-1)} \Delta t + \alpha \vec{P}_{(t-1)} \Delta t \quad (13)$$

提案システムにおける対話の位置づけ 提案システムにおける対話の位置づけを以下、図 54 に示す。ユーザは仕事を始めたとする。そうした場合、第一にファイルマネージャとの対話により仕事場を作成する。その仕事場に情報が足りないとき Mason の類似によるデータの集合と、MALLET による確立的リストのデータを見る。同時に類似と Mason を基とした類似と確率を足し合わせたデー

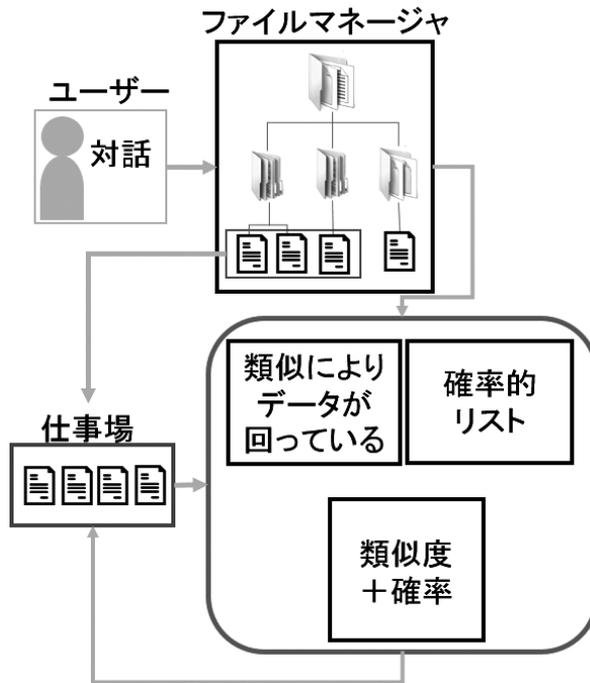


図 54 対話

タの集合の確認を行う。

その中でユーザが自分の必要なデータが全て揃ったなら終了、まだ足りないと感じたなら、上記で行った三つの項目の中にあつた必要ファイルを仕事場へ移動させ、同じ行動をする。そのサイクルによって自分のファイルマネージャの中にある、必要ファイルを集めることができる。

上記のファイルマネージャ、仕事場、MALLET (トピックモデル)、Mason (群知能)、類似+確率のフローチャートを以下図 55 に示す。

7.3.2 まとめ

本研究では、今の自分の仕事場に関連するファイルをファイルシステムから群知能によって集め、さらにトピックモデルによって、自分の今の仕事に関係するトピックが抽出されて、群知能の群れに集まる。トピックに関連するファイルを引き寄せることができるため、多様な関連ファイルが身近に集まることを期待され、それによってイノベーションが促されることが期待される。

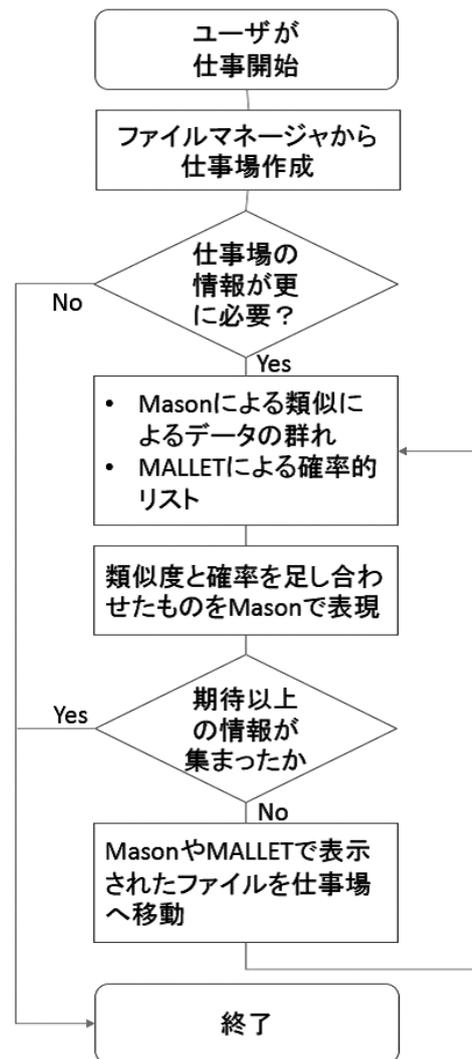


図 55 フローチャート

7.4 ケーススタディ

7.4.1 ユーザの意思 1

本章においてユーザは自分がプレイしたことのあるゲームについての資料や関連した記事を集めようとしていた。アプリケーションを使わずユーザが必要ファイルを集めるまでの過程を以下に記す。

既存のファイルマネージャの役割 ファイルマネージャは図 56 のテキストデータ 100 個のみとする。ユーザはそのテキストファイルデータの内容を殆ど理解していないものとする（ファイルの内容は Wikipedia（2016 年 1 月 31 日）の内容を引用し、概要部分からテキストファイルデータを作成）。それによってユーザは内容を見て理解した上でないと、ファイルを集めることができない。テキストの名前は Wikipedia で調べたページのタイトルを引用している。テキスト内の文章の内容は、調べたものの概要、登場キャラクター、制作会社、ゲームジャンル、歴史、成果などが入っているものとする。

ファイルマネージャ内では元々ジャンル別に分類されている。①～⑩のジャンルについて記す。

①ゲーム機②Nintendo ゲームソフト③Play Station ゲームソフト④その他 ゲームソフト⑤マリオキャラクター⑥アニメキャラクター⑦ゲームソフト制作会社⑧人物⑨ゲームジャンル⑩その他ただしテキストデータを分類する際、ユーザが直観的に入れたもので、時と場合によっては見方が変わりジャンルが変わることもある。

仕事場の役割 ユーザはファイルマネージャ内から自分のプレイしたことのあるゲームについて、テ



図 56 ファイルマネージャ (例)

キストファイルの名前から判断し仕事場へ移動させた。

図 57 のように仕事場にユーザが好きなゲームについての記事を入れている。その中に入っているのはユーザがプレイしたことのある様々なゲーム機のゲームソフトを 10 個入れてあるものとする。ユーザはそのゲームについて関連情報や似た情報を集めたいがユーザ自身の手作業では見つけることが困難である。

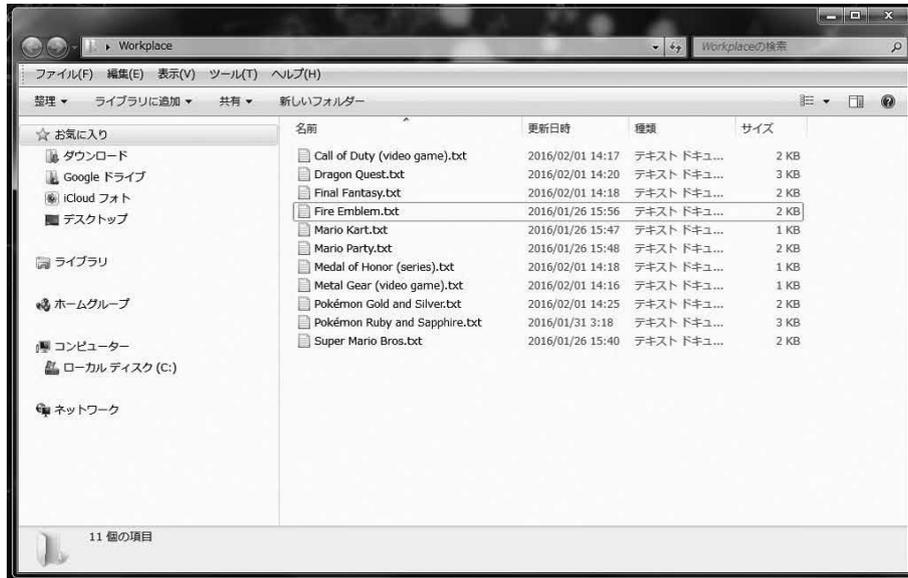


図 57 仕事場 (例)

7.4.2 ユーザの意思 2

必要ファイルが集まらなかった場合、ユーザの取るべき行動を以下に記す。また、アプリケーションの実行は同時に行っているとし、その三つの中の結果で必要とされるファイルは、ユーザ自身の意志で決める。

Mason の役割 ユーザはこのファイルマネージャの中から関連ファイルを探すためにまず Mason (群知能) と対話する必要がある。この例の中での Mason はテキストファイルの名前と内容について単語別に局所相対的に類似を判断し集まる。そして類似度によりデータが回っておりアプリケーションを止めて、ドラッグアンドドロップでファイルマネージャへ移動させることで、ファイルのショートカットを確認することができる。その中に関連ファイルがあったら仕事場へ移動させる。

本章の場合、仕事場が重要ファイルとし、類似したファイルを集める時、名前の類似では、シリーズ物ゲームで同じ名前が入ったものなどが集まり、内容では、ゲームジャンル、制作会社、登場キャラクターなどのデータから集まる。つまり名前が似ているかつ、ゲームジャンル、制作会社、登場キャラクターが似ているファイルがより重要とされる。

MALLET の役割 ユーザは MALLET を使用する。MALLET は指定したデータについて隠れたトピックについての確率を計算してコマンドプロンプト上に確率リストとして表示することができ

る。本研究での MALLET はテキストファイルの名前と内容について局所相対的、または全体について確率を計算。例の中で使用しているファイルマネージャの中には、ゲーム、Nintendo、PlayStation に関連した情報が多く入っているため、トピックもその内のどれかである。そのトピックとの確率を計算した場合、Nintendo のゲームと Playstation のゲームがより高い確率となる。

また仕事場のみを MALLET で計算させトピックを抽出し、そのトピックからファイルマネージャについての確率を計算する場合では違う結果が出てくることになる。仕事場の中には、Nintendo のゲームのテキストデータが多いことから、ファイルマネージャ全体のトピックとは違い Nintendo のゲームのみに関連したファイルを抽出できる。

上記のように二つの視点から MALLET によって確率を計算することでユーザはより多くの必要な情報を集めることができたことになる。

Mason (類似) + MALLET (確率) の役割 Mason でゲーム名、ゲームジャンル、制作会社、登場キャラクターが似ている物が集まり、MALLET ではゲーム会社二つのゲームソフトが高い確率で集まる。その二つを足し合わせることでゲームの名前が似ていなくても、ゲーム会社が同じであれば集まってくることになる。ユーザはゲストなどで自分の好きなゲームのキャラクターが出ているゲームなど新しいゲームを発見できたこととなる。つまりベイズ主義と頻度主義^[83]の間に生じるズレにより思わぬ発見をユーザはしたことになる。

期待通りでなかった場合 上記 4.1.1~4.1.1 の中で、ユーザが期待通りの結果が出なかった場合は、その中で自分が必要だと感じたファイルを仕事場へ移動させ同じ行動をすることで、また違う結果が出る。その行動の繰り返しの中で自分の必要を集めることができるシステムである。

7.4.3 ケーススタディまとめ

この一例の中では思いもよらなかったファイルとは、自分がやったことのあるゲームのキャラクターがゲスト出演などしているゲームであり、関連した情報を集めるつもりが、新しい発見につながり、やってみたいゲームまでをも見つけることができた。

またケーススタディのようにユーザにとって良い結果が出た場合、それを仕事場に移動させ、ファイルマネージャに情報を増やす行動を取ることでより良い結果を導くことができる。

7.5 結論

7.5.1 まとめ

仕事場をベースに考え、Mason により類似度によって集め、MALLET によって確立的にファイルを集める。その結果からファイルマネージャから自分の仕事場を変化させていく。ユーザが必要なファイルを仕事場へ移動させる。ファイルマネージャ、仕事場、Mason、MALLET の四つのシステムとユーザとの対話の中で思いもよらないファイルが出てくることがある。

8 その他の研究成果

8.1 民具データベースのRDF化とオントロジーを導入した情報検索システム [目的 1]

近年研究や教育に役立てることを目的として、収蔵資料をデータベース化してウェブ上で公開する博物館が増加している。しかしこの博物館資料のデータベースが関係データベースに代表される従来の技術で実現される場合、資料の分類の変更や資料について熟知していない者の利用に十分に対応することはできない。また現状では博物館によって資料の整理の仕方が異なるため、複数の博物館のデータベースの情報を統合したり横断的に検索したりすることは極めて困難である。以上の問題を改善する試みとして、本研究では福島県只見町の民具データベースを対象にオントロジーを導入した情報検索システムを提案する。提案システムではデータベースはRDF化され、その上でデータの意味を扱うのに必要となる語彙と知識がOWLによってオントロジーに記述されている。こうすることでデータベースに記載された資料に関するメタデータをコンピュータが理解できるようになるため、意味検索や他のデータベースとの相互運用の簡易化が可能になる。

8.2 リレーショナルモデルによるデジタルアーカイブのための民具データベースの構築 [目的 1]

近年では資料情報のデジタル化が進みインターネット上で資料を検索し、参照可能な施設も増えつつある。しかし、国内でのデジタル化される資料情報について保存方針は未整備であり統一されていない。そのため、情報を共有するにあたって互換性の確保は大きな課題である。加えて、博物館の資料は多様であり、デジタル化において各博物館によって、あるいは資料群によって異なる規格で作成している。特に民俗学の分野では、研究者によって資料の分類方法は一様ではない。また、名称に方言を含むことが多く、類義語から記載のずれが生じやすいという問題がある。本研究では、民俗資料特有の情報の維持と資料情報の互換性の確保を目的とし、民俗資料の一例として福島県南会津郡只見町に伝わる民具を対象とする。民具の持つ重要な要素に注目し、民具情報構造化モデルを提案する。また、民具のデータベース化を行うとともに、民具の検索システムについても検討を行った。

8.3 自己組織化可能な群知能を用いた情報リソース管理 [目的 1]

ビッグデータやクラウドで知られている様々な情報データは、ファイルの持つ複数の属性により、管理が複雑になる。また、現在は木構造がファイル管理の主流であるが、直感的にファイルの位置を把握することが難しい。一方で、多数のファイルであっても、類似性の高いファイルごとに大きく分類されていれば、直観的にユーザがデータを扱えると考えられる。そこで本研究では、各データを、自己組織化可能な群知能データと見なし、それらの群知能データを制御することで、視覚的にデータを管理可能にする。提案管理手法は、データ管理の過程に応じて自己組織化する手法にも適用可能である。

8.4 ACOを用いた検索過程を重視した検索手法 [目的 1]

インターネットで検索をするときに、比較的簡単に一定レベルの検索結果は得ることができる。しかし、検索の質は常に十分とは限らない。もしその情報についてさらに深く理解しようとするれば、関

連した情報についても調査する必要がある。そのとき、その分野に詳しい人の検索結果やその過程を参考にできれば、より詳細で多くの情報を検索過程から得ることができる。本稿では、検索過程の質を考慮した検索手法を提案する。群知能の ACO を用いた手法と「推薦システム」と呼ばれるシステムを用いた手法の二つを提案する。

8.5 半脆弱性電子透かしによる印刷物の複製検知と認証 [目的 2]

QR コードは低コストで大容量の情報を格納することができ、その情報を誰でも取得することができる。近年では航空券の自動チェックインやビットコインの認証などに利用されており、その需要は拡大している。しかし、紙コンテンツはコピー機やスキャナなどによってデジタルコンテンツへと変換を行うことが簡単である。そのため誰でもコンテンツの改ざんといった著作権侵害が発生しやすく、特に複製による金銭的損失は社会問題になりつつある。このようなリスクに対応するためにデジタルコンテンツ上で対策を行い著作権の保護を目的とする電子透かしという技術がある。電子透かしとは画像や音声などのメディアコンテンツに人間が知覚できない形で微小な変更を行い、情報を埋め込むことおよび抽出する技術のことである。しかしながら電子透かしはデジタルデータに対応するための研究が多くされているが、印刷物に適用するあまりに少ない。先行研究では DCT や DWT を用いて電子透かしを行うことで半脆弱性を実現している。本研究ではメディアフィルタの平滑化によって複製時におけるインパルス性雑音の強調と DWT を用いることでブロック歪みやノイズに対して耐性を持たせる。また実社会で使用することを想定し、真贋判定を透かし情報である QR コードの読み取りの可否にすることで、判定にオリジナル透かし画像が不要となり、人間の手によらない機械的な判別が実現した。

8.6 RBAC とハイパーグラフを用いた推論攻撃に対する個人情報保護 [目的 2]

一般に行政が扱うデータは膨大で、さらにマイナンバー制度では市役所と市役所、または部署と部署の間に膨大なデータが行き来することとなる。一つの部署で扱っていたデータが複数の部署間で扱われることとなると、少量だったデータが膨大になりデータ移行の際に改ざんや情報漏えいが起こる可能性がある。また、利用者には次のような懸念があるとされる。

- 情報を取られたくない
- 追跡されたくない
- 都合の悪い情報は忘れてほしい

つまり個人情報の流出を恐れて情報を残したくないのである。さらに集積・集約された個人情報を基に推論によって、流出元の知人である特定の個人の個人情報が暴かれてしまうといった懸念がある。さらに複数のデータが一つに集まった際の処理が複雑になることも考えられる。マイナンバー制度ではマイナンバーに所得情報、納税実績、社会保障などの情報を紐づけて手続きを簡略化することを目的としている。紐づけることによってマイナンバーが漏えいした時にそのものの様々な個人情報の漏えいにつながる危険性は否定できない。具体的には、

- 誰がどのように利用するのかわからない
- 目的不明確な名称により軽々しい扱いを助長してしまう

- 高度な情報収集により情報への勝手な意味づけがされてしまう

などの危険性がある。本研究ではハイパーグラフによる推論経路分析にロールベースアクセス制御モデルの「役割」という主体を制約条件として付加することで、アクセス権限管理の効率化、推論経路の削減ができると考え、ハイパーグラフによる推論経路分析を主体（役割）と客体の両面から評価するセキュリティモデルを提案した。

8.7 シグネチャ型IDSと機械学習手法を組み合わせた侵入検知システムの提案 [目的 2]

近年、企業や国家を標的としたサイバーセキュリティ事件や事故が社会問題となり、標的とされた企業や国家は深刻な被害を受けている。^[1]警視庁の調べによると、警視庁が観測したインターネット上の不審なアクセスは、2014年で1日1IPアドレス当たり684.9件と、前年より52.8%増加していることが分かっている。また、2015年上半期の標的型メール攻撃の認知件数も1,472件と、前年同期と比べて1,267件、618%増加している。2015年5月には、日本年金機構の年金情報管理システムサーバが、外部からの不正アクセスによって個人情報が出てしまう事件が起こった。さらに2015年はインターネットバンキングに係る不正送金事犯の被害が、過去最悪を記録した2014年をさらに上回る記録を更新した。こうしたことから、ますます巧妙化する不正アクセスの検知はネットワークセキュリティでの重要な課題として挙げることができるだろう。マルウェアや不正侵入などのサイバー攻撃に対応するセキュリティ技術の一つに侵入検知システム（Intrusion Detection System：IDS）がある。IDSは、セキュリティを確保したいネットワークあるいはサーバへの経路上を監視し、パケットやデータストリームなどの検査を行い、セキュリティを侵害する恐れのあるものを検知し、管理者へ通知する。その中でも現在一般的に使用されているのがシグネチャ型IDSだ。特徴としては、既知の攻撃パターンをデータベースに登録しておき、サンプリングしたパケットデータに対して、パターンマッチングを行うことで攻撃の検知を行う。しかし、シグネチャ型IDSでは既知の攻撃を検知することはできても日々増え続ける新種や亜種の攻撃を検知することはできない。そこで現在注目されているのがアノマリ型IDSである。アノマリ型IDSはネットワーク上のパケットとIDSに登録されている特徴との類似性を使って攻撃の検知を行う。登録されているパターンとパケットの類似性から判断するため、そのパケットが登録されたパターンと完全に一致していなくても検知が可能になっている。そのためアノマリ型IDSは、新種や亜種の攻撃パターンにも対応できるIDSだといえる。しかし、類似しているパケットを検出するがゆえにアノマリ型IDSは誤検知が多くなるという問題点が発生しやすい。人間があらかじめIDSに新しくかつ膨大な判別規則を知識として与える事も考えられるが、ネットワークやサーバへの攻撃は多種多様であり限界がある。そこで、機械学習を用いて過去の経験から機械が自動的に判別規則を生成するIDSの研究が進められている。既に挙げた通り、誤検知が多いとされているアノマリ型IDSは、判別の精度や再学習の速度がより重要視される。先行研究では、機械学習手法であるSOM（自己組織化マップ）と決定木を組み合わせたIDSを提案し、逐次学習が可能な自己組織化マップと、一定量のまとまったデータで定期的に学習する決定木を組み合わせる事で徐々に変化する攻撃通信を検知し続ける事を目標にしている。システム全体として90%後半の正答率を出すことに成功しているが、他の論文などでも採用されている目標は99%、誤検知（正常通信を攻撃通信と判断してしまう確率）の目標は0.01%であることから、より確

実性の高い検知を行うシステムが求められる。また機械学習手法を用いる場合、実環境における学習データの入手が困難といった問題も挙げられている。

問題点でも挙げた通り、IDS は高い精度で確実な検知を行うことが第一に求められる。それに加え日々増え続ける新種や亜種の不正アクセス手法にも対応できる IDS が望ましい。そこで本研究の目的は、既知の攻撃の検知率が高いシグネチャ型の確実な検知精度を確保するとともに、決定木を用いて未知の攻撃の検知でも対応できるような IDS を提案した。今回、数ある機械学習手法の中で決定木を選択した理由としては、決定木は分類速度が高速で、なぜそのような結果になったのかを解析しやすいといった利点があるため、シグネチャ型に決定木の性質を組み合わせてもシグネチャ型単体で行う処理時間を大きく変化させることなく未知の攻撃にも対応できることを示した。

8.8 群知能を用いた動的なファイルマネージャの提案 [目的 4]

近年、スマートフォン、タブレット端末、PC など個人の持つ電子端末の量が増えている。それに準じ、一人の人間が管理する情報やファイルの数も増してきている。ファイルの数が増えると管理の手間が増し、扱いにくくなっていく。さらにクラウドやビッグデータといった、大量のファイルを扱うシステムも多く開発され、管理は複雑さを増す一方である。人間が情報をより扱いやすくするためには、より人間が体感的に扱いやすいファイルシステム、およびそのユーザインターフェイスが必要となっている。ファイルの多次元の空間的な管理は、システム上要素が多い分コンピュータでは管理がしやすいが、人間が視覚的に把握するのは難しい。そこで、人間がファイルを視覚的により見やすいファイルシステムが必要となる。さらにクラウドやビッグデータといった、大量のファイルを扱うシステムも多く開発され、管理は複雑さを増す一方である上に、個人や企業が持つ PC やスマートフォンなどの電子端末があり、これらにはファイルが入っている。さらに近年では、ネットワークの普及から外部サーバへファイルを預け、またその外部サーバのサービスを基に個人や企業のデータを管理していることも少なくない。

このように情報やデータが端末で分散しているにもかかわらず、これらの情報を管理するためには、各端末で木構造のファイルを管理する必要があるため、これらのデータの場所を把握することは難しいこれは一つのファイルが、フォルダという一つの属性しか持たないために引き起こされる、いわゆる“こうもり問題”が存在しているためである。“こうもり問題”とはこうもりが獣か鳥に属するのか分からず、正しく分類できないという問題であり、これは様々な属性を持ち合わせる今日のファイルのフォルダ分け問題と等しい。また、視覚的な面で欲しいデータがどこにあるのか直観的に分かりにくく、検索に時間がかかるため、ユーザ個人や組織の活動を少なからず阻害している。

本研究では、「相互類似による引力と斥力を表現した群知能を用いた情報リソースの管理」を参考にしている。この研究では、データの分野やジャンルなどの類似度に応じてファイルを集めることで、クラウド上や個人の持つ端末の中にある多くのファイルを一目で把握可能とするファイルマネージャが提案されている。そしてそのファイルを集める手法に群知能の一種である boid が使用されている。しかし boid は本来動き続け群れを形成することを目的をしているため、ファイルマネージャに適用させるためには、パーティクルの動きを止めることを考慮し計算をしなくてはならない。そこで本研究では、類似度に応じて集める過程をパーティクルの配置の最適化問題と見なすことで、パー

ディクルの動きを考慮しなくてよいファイルマネージャの提案を行う。本研究では、ファイルをただ整理するのではなく、ユーザの振る舞いにおいて使用しているファイルを集めることで行う。この集める手法に群知能を導入し、ファイル自身が意図を持って群れるようにする。その一連の動作をファイルの振る舞いと呼ぶ。そうすることでユーザが気づかなかったファイル同士の関係性に気づけ、視覚的にも見やすくなる。本論文では、「相互類似による引力と斥力を表現した群知能を用いた情報リソースの管理」（平成 26 年度神奈川大学修士論文）の問題点に焦点を当てる。

9 今後の課題と展望

9.1 自己点検・評価

第三期共同研究では、インターネット・エコミュージアムや只見町に開設予定の民俗博物館において必要なデータマイニングやデータの入力や検索に適したユーザインタフェースなどの基盤技術を開発することが目的であった。

オントロジーを用いたデータマイニングの実際の資料に対する応用については、基盤技術の開発は当初の目標をほぼ達成している。実際の只見カードのデータでの運用は、関係データベースで表現されている部分について RDF で表現した XML データベースへの変換が必須であるが、試験的に一部の変換でのみ評価を行っているため、全てのデータについて変換を行うことが、残された課題となる。また、只見カードに記載されてる非定形の情報は関係データベース化されていないため、これをどのようにデータベース化していくかも重要な課題となる。

資料のデータベースをクラウド化する際の個人情報保護と著作権管理については、情報漏えいを検出するカバートチャンネルのモデルや著作権管理のための電子透かし、自律的な著作権管理を実現するエージェントベースの情報カプセルなど基盤技術の開発は目標を達成しているが、実用に耐える実装が残された課題である。また、個人情報保護や著作権管理へのブロックチェーンの適用などが今後の課題となる。

資料の整理とデータ入力や流通を円滑に行うためのゲーム理論や群知能などに基づく価値交換モデルの構築については、ゲーム理論に基づく手法は基本的ないくつかのモデルを提案できたが、実際の研究者のコミュニティにおいてどのモデルが最適か検討が必要となる。また、これを実用に供するように実装することも今後の課題となる。群知能に基づく手法については、構想にとどまり十分に検討することができなかった。

資料を直観的に取り扱うことを可能とする、群知能を用いた操作のコンテキストに基づくユーザインタフェースの構築については、いくつか理論的な検討を行い基礎的なフレームワークを構築し、初歩的なシミュレーションや実装を行ったが、実用システムを構築して評価実験などを行うことが今後の課題である。

9.2 第四期事業計画

第三期の研究成果を踏まえて、「非文字資料研究のコミュニティにおける知識とサービスの効率的な検索と安全安心な流通」というテーマで研究を行う。

1. **知識とサービス、物の流通と価値交換** 非文字資料を研究者間および一般ユーザとの間で知識、サービス、資料をやり取りするためにゲーム理論によりモデル化を行う。また、ビットコインで注目されている自立分散的に事象や価値の移転を記録するブロックチェーンの技術を用いて安全安心な価値交換を行うシステムを構築する。
2. **知識とサービスの検索とマイニング** 非文字資料のデータベースや研究者が資料を検索する際に、作業の流れであるコンテキストの一面に着目して情報の類似度などに基づいてファイルの自己組織化などによりユーザに対して最適な情報を提示する。
3. **個人情報や重要情報、著作権の管理** ブロックチェーンを用いたアクセス履歴管理を行うことで個人情報に対するハイパーグラフを用いた推論攻撃検出の実装を容易にする。また、ブロックチェーンによる価値移転により著作権管理を効率的に行う方法を提案する。

参考文献

- [1] 吉岡信和、田原康之、本位田真一、“モバイルエージェントによる柔軟なコンテンツ流通を実現するアクティブコンテンツ”、情報処理学会論文誌 Vol. 44, No. SIG18 (TOD20), pp. 45-57, Dec. 2003.
- [2] 阿部剛仁、谷口展郎、森賀邦広、塩野入理、櫻井紀彦、“メディアフレームワークに基づくコンテンツ保護処理方式と自律制御型情報カプセルの実現”、情報処理学会論文誌 Vol. 44, No. SIG12 (TOD19), pp. 64-73, Sep. 2003.
- [3] 加賀美千春、森賀邦広、塩野入理、櫻井紀彦、“コンテンツ流通における自律管理を目的としたカプセル化コンテンツ Matryoshka”、NTT サイバースペース研究所、マルチメディア通信と分散処理 97-18、コンピュータセキュリティ 8-18, Mar. 2000.
- [4] 谷口展郎、難波功次、塩野入理、“情報カプセル流通における利用者システム保護”、NTT サイバースペース研究所、Feb. 2001.
- [5] 神崎正英、“Dublin Core (ダブリン・コア)：ウェブ資源メタデータの共通語彙”、
<http://www.kanzaki.com/docs/sw/dublin-core.html> Feb.2016.
- [6] 山田祥寛、10日で覚えるXML入門教室第2版、翔泳社、2004.
- [7] 櫻井紀彦、“カプセル化コンテンツの動向と展望”、NTT サイバースペース研究所、電子化知的財産社会基盤、12-1, June. 1. 2001.
- [8] 石川冬樹、吉岡信和、本位田真一、“Event Calculusに基づく合意を用いたモバイルエージェントの協調”、電子情報通信学会論文誌 D, Vol. J90-D, No. 9, pp. 2,349-2,364, Sep. 2007.
- [9] F. Ishikawa, N. Yoshioka, S. Honiden, “Smartive: Agreement-based Mobile Composite Agents for Multimedia Services,” IEEE, International Conference on Computational Intelligence for Mod-ellingControl and Automation, and International Conference on Intelligent Agents, Web Tech-nologies and Internet Commerce, 0-7695-2731-0/06, 2006.
- [10] 本位田真一、吉岡信和、石川冬樹、“自由でかつ安全なコンテンツ流通を実現するエージェントフレームワーク”、2006。
http://www.soumu.go.jp/main_sosiki/joho_tsusin/scope/event/h19yokousyu/session6/sangakukan5.pdf,
参照 Feb. 2, 2016.
- [11] 星野寛、山田篤、鎌田浩典、“コンテンツ複合的権利記述による権利保護と流通支援”、京都高度技術研究所、電子化知的財産・社会基盤 2-1, Sep. 1998.
- [12] 熊沢雅之、鎌田浩典、山田篤、星野寛、上林彌彦 “多権利者間の権利関係及び利益分配方式の記述によるコンテンツ再利用支援”、京都高度技術研究所、電子化知的財産社会基盤、3-9, pp. 65-72, Jan. 30. 1999.

- [13] 乙黒泰輝、“Dublin Core を用いた著作権管理”、神奈川大学工学部電子フロンティア学科 2010 年度卒業論文。
- [14] 山田孔太、木下宏揚、森住哲也、“情報フィルタと情報カプセルによる著作権・所有権保護システム”、情報処理学会研究報告、2006-CSEC-34 (7), 20. July. 2006。
- [15] 玉井誠、庵祥子 “情報販売における不正コピー防止方式の実装”、NTT ソフトウェア研究所、情報処理学会第 57 回全国大会論文集、3-25-26, 1998。
- [16] “THE APACHE SOFTWARE FOUNDATION,” “the Apache Xerces,”
<http://xerces.apache.org/>参照 Feb. 2, 2016。
- [17] 著作権のひろば管理人ひの “著作権の歴史”、<http://cozylaw.com/copy/tyosakuken/history.htm> 参照 Feb. 2, 2016。
- [18] Microsoft, “Microsoft Windows Media デジタル著作権管理 (DRM)”、
<http://www.microsoft.com/japan/windows/windowsmedia/drm/default.aspx> 参照 Feb. 2, 2016。
- [19] Incept Inc., “IT 用語辞典 e-Words XML [Extensible Markup Language)”,
<http://e-words.jp/w/XML.html>, 参照 Feb. 2, 2016。
- [20] シャロン・ザクワフ、スコット・ホンメル、ジャコブ・ロイヤル、Java チュートリアル第 4 版、安藤慶一、株式会社ピアソンエデュケーション、2007。
- [21] 鈴木遼、“推論による情報漏えい防止のためのオブジェクト関係の視覚化”、神奈川大学 2009 年度卒業論文、神奈川、Mar. 2009。
- [22] 総務省、“マイナンバー制度”、
http://www.soumu.go.jp/kojinbango_card/01.html, 参照 Oct. 2015。
- [23] 鈴木遼、“推論による情報漏えい防止のためのハイパーグラフによる依存関係のモデル化とアルゴリズム”、神奈川大学 2011 年度修士論文、神奈川、Mar. 2011。
- [24] 酒井剛典、森住哲也、畔上昭司、小松充史、稲積泰宏、木下宏揚、“Covert Channel 分析メカニズムと EJB による情報フィルタの構築”、2006 年暗号と情報セキュリティシンポジウム、2006。
- [25] 鈴木遼、鈴木一弘、森住哲也、木下宏揚、“推論による情報漏えい防止のためのハイパーグラフモデル”、電子情報通信学会論文誌 D 情報・システム、J95-D (4), pp. 812-824, 2012。
- [26] 内野雄策、“SNS における情報漏洩を防止するための情報フィルタの適用”、神奈川大学 2008 年度卒業論文、神奈川、Mar. 2008。
- [27] David F. Ferraiolo, D. Richard Kuhn, “Role-Based Access Controls”, 15th National Computer Security Conference (1992), Baltimore, pp. 554-563, Oct. 13-16, 1992。
- [28] 横浜市総務局人事部人事課、“横浜市組織図”、
<http://www.city.yokohama.lg.jp/somu/org/jinji/soshikizu/soshikizu.pdf>, May. 2015。
- [29] 伊藤圭祐、“マイナンバー制度におけるアクセス制御”、神奈川大学 2013 年度卒業論文、神奈川、May. 2013。
- [30] 境頭宏、鶴林尚靖、中島震、“RBAC モデルの形式検証と修正支援”、九州工業大学情報工学部、国立情報学研究所、情報処理学会研究報告ソフトウェア工学 (SE)、no. 29, pp. 49-56, Mar. 17, 2008。
- [31] 鈴木大輝、小林秀幸、岡本圭史、高橋薫、“RBAC モデルの形式検証”、仙台高等専門学校、情報科学技術フォーラム講演論文集、vol. 11, no. 1, pp. 221-222, Sep. 4, 2012。
- [32] 平川学、“著作権保護における電子透かしサービスモデルの有効性検証”、東京工業大学、大学院社会理工学研究科、経営工学専攻、2008 年度修士論文。
- [33] 匂坂直樹、“QR コードを用いたウェブレット変換による電子透かし”、神奈川大学、2013 年度卒業論文。
- [34] 佐藤貴将、“電子透かしの再撮への耐性”、神奈川大学、2014 年度卒業論文。
- [35] 金城和志、稲葉宏幸、“幾何学変換に耐性を有する線画用電子透かしの提案”、情報処理学会論文誌 Vol. 46, No. 8, pp. 1,834-1,842, Aug. 2005。

- [36] 松原良平、“周波数変換による静止画像への情報埋め込みに関する一考察”、早稲田大学、理工学部、情報学科、2004年度卒業論文。
- [37] 荒木研二、“証明書の顔写真への電子透かしによるなりすまし防止”、神奈川大学、2011年度卒業論文。
- [38] The Mathworks, Inc. “MathWorks 日本”、<http://jp.mathworks.com/help/images/ref/hough.html>、2015。
- [39] 磯崎邦隆、“電子透かしにおけるマスキング効果の主観評価”、東海大学、電子情報学部、情報メディア学科、2006年度卒業論文。
- [40] “デジタル画像処理”、CG-ARTS 協会、2004。
- [41] 柴田且崇、沢田克敏、中村栄治、“DCT および VQ を用いた画像電子透かし”、愛知工業大学研究報告 Vol. 47、pp. 237-245、2012年。
- [42] 水本匡、松井甲子雄、“DCT を用いた電子透かしの印刷取り込み耐性の検討”、電子情報通信学会論文誌、Vol. J85-A、No. 4 pp. 451-459、Apr. 2002。
- [43] 泉明宏、杉岡謙太、浜野大輔、“画像伸縮に DCT を用いた解像度変換”、岡山理科大学、総合情報学部、情報科学科、2008年度卒業論文。
- [44] 井上尚、宮崎明雄、島津幹夫、桂卓史、“ウェーブレット変換を用いた画像信号に関する電子透かし方式”、映像情報メディア学会誌 Vol. 52, No. 12, 1998。
- [45] The Mathworks, Inc. “MathWorks 日本”、<http://jp.mathworks.com/>、2015。
- [46] 濱田貴広、川村秀憲、山本雅人、大内東、“適応的ハイパーゲームによる市場取引のモデル化と構造分析”、信学技法、2002年、1月、pp. 33-40 (2002)
- [47] 武藤正義、“利他的効用関数による協力的秩序形成の可能性 —— 進化ゲーム理論的アプローチ ——”、東京工業大学 2000年度修士論文、2001年、1月 (2001)
- [48] 武藤正義、“日常世界的秩序問題のゲーム理論的分析 —— 《二重の選択》を手がかりに ——”、2005年 (2005)
- [49] 久野木彩子、山名早人、“繰り返し囚人のジレンマゲームのネットオークションへの適応と協調行動と落札の関係の考察” 早稲田大学、2004。
- [50] 横山智生、“繰り返し囚人のジレンマゲームを対象とした競合共進化モデルに関する研究”、2001年、2月 (2001)
- [51] 久慈利武、“秩序問題への個人主義アプローチの可能性 —— 規範主義社会学と契約理論からの挑戦との戦い ——” 理論と方法 Vol. 6、1991。
- [52] 武藤正義、“日常世界における「囚人のジレンマ」の意味”、2005年、10月 (2005)
- [53] 平山貴之、“市場取引における価値交換システム”、2014年、2月 (2014)
- [54] “Genetic Algorithm Research and Applications Group”、2008年 (2008)
- [55] 八十二文化財団、“日本の貨幣の歴史”、2008年 (2008)
- [56] 清水孝治、戸田瑛人、木下宏揚、森住哲也、“人間関係ダイアグラム評価を導入した地域 SNS での地域通貨の使用”、信学技報、vol. 108、no. 331、SITE2008-37、pp. 7-12、2008年12月 (2008)
- [57] 三浦一輝、“「地域通貨」の流通に関する理論分析”、法政大学大学院紀要、60、pp. 69-76、2008年3月 (2008)
- [58] 岩田学、秋山英三、“相手の「協力度」を考慮する戦略と、ネットワーク上での協力の進化”、第8回情報科学技術フォーラム、2009年 (2009)
- [59] 寺岡義伸、北篠仁志、“価値に変動を伴う売り出しのサイレント・ゲーム”、日本オペレーションズ・リサーチ学会、2006年 (2006)
- [60] 寺岡義伸、山田康吉、“価値が時間に関係する縄張りのゲーム”、日本オペレーションズ・リサーチ学会、1996年 (1996)
- [61] 高橋真吾、“ハイパーゲーム分析”、2001年、2月 (2001)

- [62] “中国の貨幣制度史”、Wikipedia
- [63] “新貨条例”、Wikipedia
- [64] 村上研二、“第2章 遺伝的アルゴリズムの処理手順”、村上・泉田研究室 HP 製作委員会、2001年(2001)
- [65] 宮元裕樹、酒居敬一、“セマンティックファイルシステムのフレームワークの提案”、研究報告計算機アーキテクチャ (ARC)、vol. 2011-ARC-194, no. 10, pp. 1-8, Mar. 2011。
- [66] 石川憲一、森嶋厚行、田島敬史、“大規模ドキュメント空間管理のための意味ファイルシステムの構築”、情報処理学会研究報告データベースシステム (DBS)、vol. 78, no 2006-DBS-140, pp. 497-503, July 2006。
- [67] 坂本貴史 “IA THINKING,” ワークスコーポレーション：初版、pp. 1-224, Mar. 2011。
- [68] 寺岡照彦、丸山稔、中村泰明、西田正吾、“空間検索を効率化した時空間データ管理構造の提案—多次元 Persistent Tree.” 信学会論文誌 (D)、vol. J78-D2, no. 9, pp. 1,346-1,355, 1995。
- [69] Abraham A, Grosan C, Ramos V, editors “Swarm intelligence in data mining,” 2006.
- [70] 神嶋敏弘、“データマイニング分野のクラスタリング手法”、人工知能学会誌、vol. 18, no. 1, pp. 59-65, Jan. 2003。
- [71] 田中克弥、“米国におけるクラウドコンピューティング事情と適用範囲への一考察”、
<http://www.rbbtoday.com/article/2010/12/26/73139.html> ユニシス技報、vol. 29, no. 4, Mar. 2010。
- [72] 堀井秀之、“社会問題解決のための知の構造化”、情報処理学会、vol. 48, no. 8, pp. 819-823, Aug. 2007。
- [73] 大崎史裕、“データベースに対するインタフェースとしてのファイルシステムの実装”、高知工科大学情報システム工学科学士学位論文、Mar. 2008。
- [74] T. Rizzo, “ntroducing the New Windows File System,”
<http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwinfs/html/winfs03112004-3-11> Microsoft Corp., 17 Mar. 2004.
- [75] 石田克憲、小泉駿、宮田純子、森住哲也、木下宏揚 “相違類似による引力と斥力を表現した群知能を用いた情報リソースの管理”、研究報告インターネットと運用技術 (IOT)、vol 2015-IOT-28, no. 39, pp. 1-6, 2015, Feb. 2015。
- [76] 山崎智史 “Boid モデルを用いたファイルシステム”、神奈川大学電子情報フロンティア学科 2014 年度卒業論文。
- [77] 岩田具治、“潜在トピックモデルを用いたデータマイニング”、NTT コミュニケーション科学基礎研究所”、電子情報通信学会誌、vol. 97, no. 5, pp. 405-409, May 2014。
- [78] GushiShow、“数式をなるべく使わずにトピックモデルの解説にチャレンジ”、
<http://qiita.com/GushiSnow/items/8156d440540b0a11dfe6>, 2015
- [79] 上田修功、山田武士、“ノンパラメトピックベイズモデル”、日本電信電話 (株) NTT コミュニケーション科学基礎研究所
- [80] McCallum, Andrew Kachites. “MALLET: A Machine Learning for Language Toolkit.”
<http://mallet.cs.umass.edu>, 2002.
- [81] Sean Luke, Gabriel Catalin Balan, Liviu Panait Keith Sullivan, Claudio Cioffi-Revilla, Sean Paus, Daniel Kuebrich, Joey Harrison, Ankur Desai “MASON,” <http://cs.gmu.edu/~eclab/projects/mason/>, 2003.
- [82] 加納寿浩、“高機能なオープンソース全文検索システム”、<https://thinkit.co.jp/free/article/0612/8/2/>、NTT データ、Dec. 2016。
- [83] 池田真治、“哲学のすすめ 統計の哲学入門”、富山大学、日本哲学会、pp. 1-6、July 2013。

主な業績

1. Rie Jimbo, Sumiko Miyata, Kazumitsu Matsuzawa, Hirotsugu Kinoshita, "RDFization of Database for Folk Implements and Ontology-assisted Information Retrieval System," IEEE ICCE 2015 Consumer Electronics—Taiwan (ICCE-TW), 2015 IEEE International Conference on Year: 2015, pp.490-491, DOI: 10.1109/ICCE-TW.2015.7217015
2. Toshihide Hanyu, Sumiko Miyata, Tetsuya Morizumi, Hirotsugu Kinoshita, "Development of the folk implements database for the digital archive," IEEE ICCE 2015 Consumer Electronics—Taiwan (ICCE-TW), 2015 IEEE International Conference on Year: 2015, pp.484-485, DOI: 10.1109/ICCE-TW.2015.7217011
3. Sumiko Miyata, Kazuhiro Suzuki, Tetsuya Morizumi and Hirotsugu Kinoshita, "Access control model for the My Number national identification program in Japan," In Proc. of IEEE COMPSAC 2014, pp.152-157, Jul. 2014.
4. 木下宏揚、"仮想通貨 Bitcoin を支える技術"、KDDI 総研 Nextcom Vol. 26 2016 Summer, pp. 24-33 (2016-6)。
5. 羽生敏英、宮田純子、森住哲也、木下宏揚、"デジタルアーカイブ作成を前提とした民具データベースの構築"、信学技報、vol. 114, no. 494, SITE2014-77, pp. 215-220, 2015 年 3 月。
6. 小泉駿、宮田純子、森住哲也、木下宏揚、"ACO を用いた検索過程および検索履歴を考慮した検索手法"、信学技報、vol. 114, no. 494, SITE2014-78, pp. 221-226, 2015 年 3 月。
7. 石田克憲、小泉駿、宮田純子、森住哲也、木下宏揚、"相互類似による引力と斥力を表現した群知能を用いた情報リソースの管理"、信学技報、vol. 114, no. 494, SITE2014-80, pp. 233-238, 2015 年 3 月。神保理恵、宮田純子、木下宏揚、松津和光、"自然言語処理におけるオントロジ応用についての考察"、人工知能学会第 47 回ことば工学研究会 SIG-LSE-B402, pp. 13-19, 2014 年 12 月。
8. 羽生敏英、宮田純子、森住哲也、木下宏揚、"Ontology を用いた非文字資料の検索"、信学技報、vol. 114, no. 25, SITE2014-4, pp. 53-58, 2014 年 5 月。