様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １８ 日現在

機関番号：２５４０３
研究種目：基盤研究(C)（一般）
研究期間：2019～2022
課題番号：１９Ｋ１２１０２

研究課題名（和文）多様な付加情報を活用したグラフ構造データに対する高性能グラフマイニング手法の開発

研究課題名（英文）Development of high-performance graph mining methods for graph structured data using various additional information

研究代表者
鈴木　祐介（Suzuki, Yusuke）

広島市立大学・情報科学研究科・助教

研究者番号：１０３９８４６４

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：本研究課題の目的は，データの持つ付加情報を用いることで，グラフ構造データに対する効率的なグラフマイニングアルゴリズムを開発することである．
本研究課題では，1変数項木パターンという同一構造が繰り返し出現するグラフ構造データの表現に適した木構造パターンを提案し，1変数項木パターンに対する効率のよいグラフマイニングアルゴリズムを提案した．また質問学習モデルや進化的学習手法を用いた．グラフ構造データからのグラフマイニング手法の開発を行った．さらに計算論的学習理論に基づき，ある種のグラフ文法システムによって定義される言語のクラスのPAC学習可能性について考察を行った．

研究成果の学術的意義や社会的意義
情報技術の発展に伴い，グラフ構造データは大規模化かつ増大している．これらの大規模なグラフ構造データからのデータマイニングには膨大な計算資源を必要とする．一般的なグラフ構造データに対する効率の良いグラフマイニングアルゴリズムの開発は困難である．
本研究課題では，データの持つ付加情報を用いることで，データの持つ情報を活かしつつ，グラフ構造に制限を加える．これにより，ある種のグラフ構造データに対する効率的なグラフマイニングアルゴリズムを提案した．本研究成果は，大規模なグラフ構造データからのデータマイニングにおける更なる知識の獲得と計算時間の削減に寄与するものである．

研究成果の概要（英文）：The purpose of this research is to development high-performance graph mining methods for graph structured data using various additional information.
In this research, we proposed one-variable term tree patterns as tree structured patterns suitable for representing a graph structure data in which the same structure appears repeatedly. And we proposed efficient graph mining algorithms for one-variable term tree patterns. Moreover, we developed graph mining algorithms for graph structured data using a query leaning model or an evolutionary learning method. Furthermore, based on computational learning theory, we considered the PAC learnability of a subclass of graph languages defined by parameterized Formal Graph Systems.

研究分野：グラフアルゴリズム

キーワード：グラフアルゴリズム　機械学習　グラフマイニング　グラフ構造データ

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

地図データ，分子化合物データ，ネットワークトラフィックなど複雑な構造を持つデータが増大している．構造を持つデータをグラフで表現することで，各要素の接続関係を特徴として捉えることができる．一方，構造を持つデータは接続関係だけでなく，方角や向き，距離，そして時系列など多種多様な付加情報を持っているが，単なるグラフで表現すると，これらの付加情報が失われてしまうという欠点もある．データマイニングに関しては国内外で盛んに研究がなされている．テキストデータなどからの情報抽出については商業的に実用化されている．

近年では，上記のような複雑な構造を持つデータから有益な知見を得ようという社会的な要求が高まっている．構造を持つデータから更なる有益な知見を得るためには，データの持つ構造に加え，多種多様な付加情報を多面的に検討することや，多量のデータを扱う必要がある．そのため，構造を持つデータに対し，その構造的特徴を表すグラフでのモデル化や，多量のデータに対する効率のよい特徴抽出等の課題の解決方法が求められている．

グラフで表現可能なデータからのデータマイニングをグラフマイニングとよぶ．グラフマイニングにおいて重要となるのが，グラフ同型問題やパターン照合問題，パターン抽出問題などを解くグラフマイニングアルゴリズムである．しかし，一般的なグラフに対するグラフ同型問題がNP 完全であることが知られているように，グラフ全体を対象とした効率のよいグラフマイニングアルゴリズムの開発は困難である．その一方，制限をつけたグラフのクラス，例えば，平面上に辺が交叉しないように描くことができる平面的グラフや，サイクルを持たないグラフである木，隣接頂点に順序が与えられた順序グラフ，などに対しては効率のよいグラフマイニングアルゴリズムが存在することが知られている．

２．研究の目的

本研究では構造を持つデータをグラフで表現する際に，各要素の接続関係だけではなく，方角や向き，距離，そして時系列などデータが持つ多種多様な付加情報に着目する．構造を持つデータを付加情報を持つグラフとして表現することで，各要素の接続関係だけでなく，より複雑な関係やデータの持つ特徴をより詳細に表現可能になる．さらに，これら付加情報をグラフに対する制限の一種と考えることで，付加情報を持つグラフに対する効率のよいグラフマイニングアルゴリズムが開発可能であると考えられる．

グラフパターンとして，構造的変数を持つグラフパターンである項グラフパターンについて研究を行ってきた．項グラフパターンが持つ構造的変数は任意の部分グラフを表現可能であり，従来のグラフパターンよりも表現力が大きいという点が特色である．項グラフパターンは，構造的変数の位置，個数，接続箇所などによって表現力が異なり，同じグラフのクラスを対象にしても，項グラフパターンの変数の条件がアルゴリズムの効率に影響を与える．

本研究では，構造を持つデータの付加情報を持つグラフ表現，それらに共通する特徴的な構造を表現する項グラフパターンの提案，そして提案した項グラフパターンに対するグラフマイニングアルゴリズムの開発を目的とする．

３．研究の方法

本研究の目的は，データの特徴を詳細に表現可能な付加情報を活用したグラフクラスの提案，それらに共通する特徴的な構造を表現する項グラフパターンの提案，項グラフパターンのグラフマイニングアルゴリズムの開発である．研究目的を達成するため以下の方法で研究を行った．

まず、構造を持つデータの多種多様な付加情報を活用したグラフ表現の提案と共通する特徴を表現する項グラフパターンの提案については、これまでの研究成果をもとにデータの持つ付加情報を十分に活用できるだけでなく，効率のよいグラフマイニングアルゴリズムが開発可能な制限を持つ新しいグラフクラスの提案を行う．さらに提案したグラフクラスに共通する特徴を表現可能な適切な項グラフパターンの提案を行う．

次に、提案した項グラフパターンに対する効率のよいグラフマイニングアルゴリズム開発について、本研究課題ではグラフマイニングアルゴリズムとして，項グラフパターンがグラフを表現するか判定するパターン照合アルゴリズムと，グラフ集合に共通する項グラフパターンを発見するパターン発見アルゴリズムの開発を行う．また提案したグラフマイニングアルゴリズムの実装と評価実験を行う．このとき，効率の良いグラフマイニングアルゴリズムの開発が行えるよう，利用する付加情報の取捨選択を行い，適切なグラフクラスの設定を行う．

また，計算学習理論に基づくグラフマイニングアルゴリズムの開発を行う．帰納推論，質問学習モデル，進化的手法などの計算論的学習理論に基づくグラフマイニングアルゴリズムの研究を進め，付加情報を持つグラフに対する効率のよいグラフマイニングアルゴリズム開発へと応用する．

４．研究成果

本研究課題で得られた研究成果について述べる．

(1) １変数項木パターンに対するグラフマイニングアルゴリズムの開発

項木パターンは順序木を対象とした項グラフパターンの一種である．項木パターンの変数には任意の順序木を代入することができる．変数への代入の際に各変数は変数ラベルによって区別され，同一の変数ラベルを持つ変数には同一の順序木を代入しなくてはならない．項木パターン中の全ての変数が互いに異なる変数ラベルを持つとき，線形項木パターンといい，線形でない項木パターンを非線形な項木パターンという．項木パターン t の変数に適切な順序木を代入して，順序木 T と同型になるとき項木パターン t と順序木 T はマッチするという．項木パターンのパターン照合問題とは，入力の項木パターン t と入力の順序木 T がマッチするか否かを判定する問題である．非線形な項木パターンのパターン照合問題は NP 完全な問題であることが知られている．頻出項木パターン発見問題とは，順序木のデータベースと閾値σに対して，データベース中の順序木でマッチするものの割合がσ以上の項木パターンをすべて発見する問題である．

１変数項木パターンとは，項木パターン中の全ての変数が同一の変数ラベルを持つ非線形な項木パターンの一種である．代入の制限より，１変数項木パターンの全ての変数には同一の順序木が代入される．図１に１変数項木パターンの例と代入の例を示す．

本研究課題では，１変数項木パターンのパターン照合問題を解く効率の良いパターン照合アルゴリズムを提案した．またパターン照合アルゴリズムの計算量の削減を行った．また頻出１変数項木パターン発見問題を解くパターン発見アルゴリズムの開発を行った．さらにパターン発見アルゴリズムの計算時間を削減するため，複数の変数ラベルの存在を許すが１種類の変数ラベルだけ繰り返し出現する制限を持つ非線形な項木パターンである 1rep 項木パターンを提案した．1rep 項木パターンのパターン照合問題を解く効率の良いパターン照合アルゴリズムを提案し，それを用いて頻出１変数項木パターン発見問題を解く高速なパターン発見アルゴリズムを提案した．また提案したグラフマイニングアルゴリズムを計算機上に実装し，評価実験を行った．さらに１変数項木パターン拡張である k 変数項木パターンのグラフマイニングアルゴリズムについても考察を行った．

１変数項木パターンはグラフ構造データ中に同一の構造が繰り返し現れるという付加情報を用いた項木パターンであり，従来の線形項木パターンでは表現できない特徴を表現可能である．また，非線形な項木パターンのパターン照合問題は NP 完全な問題であるが，変数の種類に制限を与えることで，１変数項木パターンのパターン照合問題は多項式時間で解くことが可能となった．将来的には１変数項木パターンの拡張である k 変数項木パターンに対するグラフマイニングアルゴリズムの開発を行うことができれば，グラフ構造データからのデータマイニングにおいて有益であると考えられる．

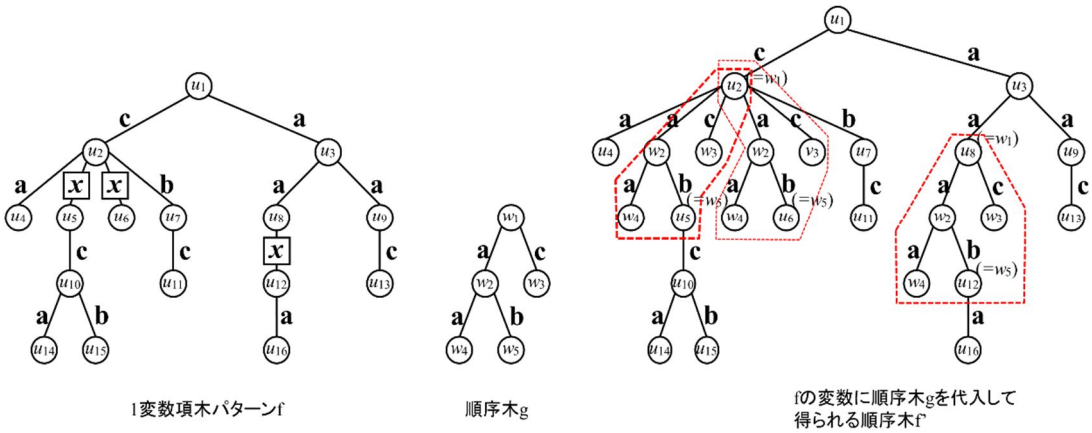

図１．１変数項木パターンの例と１変数項木パターンへの代入の例

(2) 計算論的学習理論に基づくグラフマイニングアルゴリズムの開発

定数記号と互いに異なる変数記号からなる文字列パターンを正則パターンという．正則パターン p の変数に定数記号列を代入して生成される全ての定数記号列の集合を p の正則パターン言語とよぶ．同様に項木パターン t の変数に順序木を代入して生成される全ての順序木の集合を t の項木パターン言語という．質問学習モデルでは，学習アルゴリズムはオラクルへ質問することで目標言語に関する情報を入手し，質問を繰り返すことで目標言語と同等の言語を生成する仮説（パターン）を出力する．質問学習モデルにおいて，正則パターン言語のクラスと線形項木パターン言語のクラスは１つの正例と，正例の長さの多項式回数の所属性質問を用いて質問学習可能であることが知られている．本研究課題では，正則パターン言語のクラスと線形項木パターン言語のクラスのサブクラスを，１つの正例と，正例の長さの線形回数の所属性質問を用いて学習する質問学習アルゴリズムを提案した．またグラフを対象とした深層学習モデルである Graph Convolution Network(GCN)を質問学習モデルにおけるオラクルとみなし，質問学習アルゴリズムを用いて，目標言語と同等の言語を生成する線形項木パターンを出力することで，GCN の予測根拠の可視化を行う手法を提案した．

タグ木パターンとは頂点ラベルと辺ラベルを持つ項木パターンの一種である．正例と負例の

順序木集合に対するタグ木パターン p の適合度は，(p とマッチする正例の割合＋p とマッチしない負例の割合)/2 で定義される．本研究課題では，タグ木パターンの頂点ラベルと辺ラベルにワイルドカードを導入し，進化的学習手法の一つである遺伝的プログラミングを用いて，正例と負例の順序木集合に対する適合度が高い特徴的なワイルドカード付タグ木パターンを発見するパターン発見アルゴリズムの開発を行った．BPO グラフパターン（ブロック保存型外平面的グラフパターン）とは外平面的グラフの構造的特徴を表現可能な項グラフパターンである．TTSP グラフパターンとは TTSP グラフの構造的特徴を表現可能な項グラフパターンである．本研究課題では，遺伝的プログラミングを用いて正例と負例の外平面的グラフ集合に対する適合度が高い特徴的なワイルドカード付 BPO グラフパターン集合を発見するパターン発見アルゴリズムの開発を行った．また正例と負例の TTSP グラフ集合に対する適合度が高い特徴的なワイルドカード付き TTSP グラフパターン集合を発見するパターン発見アルゴリズムの開発を行った．

(3) 物語文からの人物相関グラフ作成手法の開発

人物相関グラフとは，個々の人間を頂点，人間関係や人物間の動作などを辺で表したグラフのことである．人物相関グラフは物語やドラマの登場人物間の関係を表すためによく使用されている．物語文などのテキストを係り受け解析すると，形態素解析して得られた文字列を頂点ラベルとして持つ頂点ラベル付き木として表現することができる．テキスト d を係り受け解析して得られる頂点ラベル付き木を d の係り受け木とよぶ．本研究課題では，テキストの係り受け木の構造に着目し，有効木パターンという係り受け木の構造的特徴を表現するパターンと，それに対応する人物相関パターンを提案した．有効木パターンに適切な代入をすることでテキストの係り受け木の部分木となるとき，その代入からテキスト中の人物名と人物間の関係や動作などの情報を抽出することができる．このようにして抽出した情報を，有効木パターンに対応した人物相関パターンに代入することで人物相関グラフを作成する．本手法の流れを図 2 に示す．

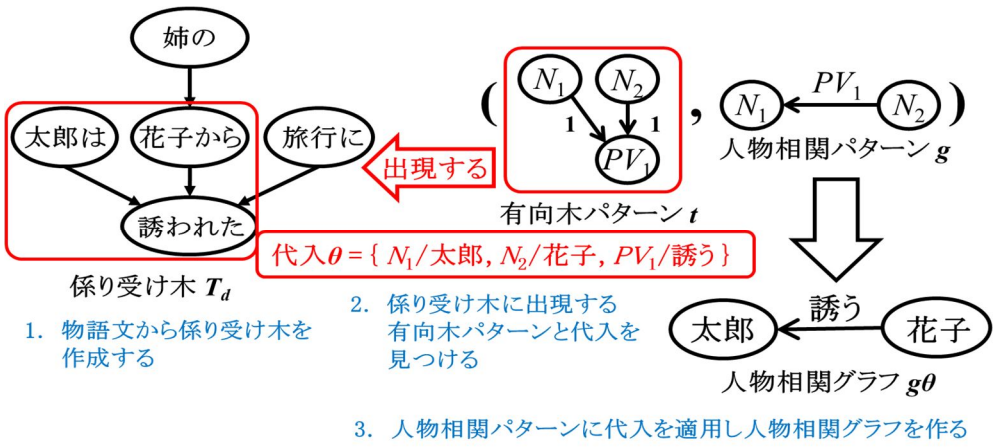

図 2．物語文からの人物相関グラフ作成手法の流れ

(4) 制限された形式グラフ体系に対する PAC 学習可能性の考察

形式グラフ体系（FGS）とは項グラフパターンを 1 階述語論理における項の代わりに扱う論理プログラミングシステムである．FGS プログラムは確定節であるグラフ書き換え規則の有限集合として定義される．FGS プログラム Γ によって生成されるグラフ集合を Γ の形式グラフ体系言語という．PAC(Probably Approximately Correct)学習は学習アルゴリズムの出力する仮説に誤差を許す計算学習理論の枠組みである．ある確率 δ と誤差 ε に対して確率 1-δ 以上で誤差が ε 以下の仮説が出力できるならば，PAC 学習可能であるという．本研究課題では，木幅定数で変数独立な遺伝的形式グラフ体系の言語においてグラフ書き換え規則のパラメータが制限されているならば多項式時間 PAC 学習可能であることを示した．図 3 に形式グラフ体系の例を示す．

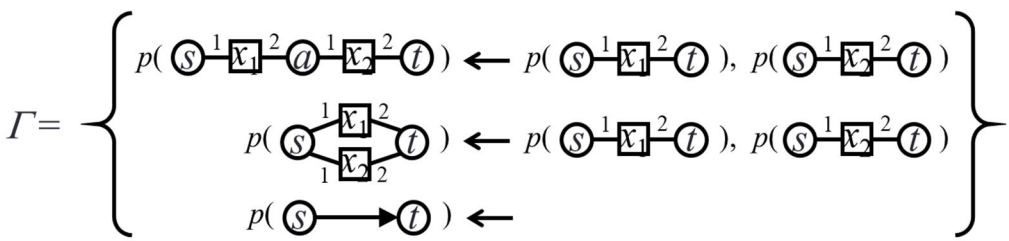

図 3．形式グラフ体系の例.

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Takayoshi SHOUDAI, Satoshi MATSUMOTO, Yusuke SUZUKI, Tomoyuki UCHIDA, Tetsuhiro MIYAHARA	４．巻 E106-A(6)
２．論文標題 Parameterized Formal Graph Systems and Their Polynomial-Time PAC learnability	５．発行年 2023年
３．雑誌名 IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences	６．最初と最後の頁 to appear
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2022EAP1052	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計17件（うち招待講演　0件／うち国際学会　2件）

１．発表者名 片山 悠，鈴木 祐介，内田 智之，宮原哲浩
２．発表標題 非線形項木パターンに対するマッチングアルゴリズムと頻出1変数項木パターン枚挙への応用
３．学会等名 2022年度「火の国情報シンポジウム2023」
４．発表年 2023年

１．発表者名 東山 的生，野口 大悟，内田 智之，正代 隆義，松本 哲志
２．発表標題 学習済超高精度GCNをオラクルとする順序項木パターンの質問学習モデルの解析と実データでの評価
３．学会等名 情報処理学会 第85回全国大会
４．発表年 2023年

１．発表者名 山本 柊一朗，宮原 哲浩，正代 隆義，鈴木 祐介，内田 智之，久保山 哲二
２．発表標題 特徴的な区間グラフパターンを獲得する進化的学習における遺伝的操作
３．学会等名 2022 IEEE SMC Hiroshima Chapter 若手研究会
４．発表年 2022年

１．発表者名
武田 直人，内田 智之，正代 隆義，松本 哲志，鈴木 祐介，宮原 哲浩

２．発表標題
線形パターンの質問学習アルゴリズムによる深層学習モデルの予測根拠の可視化

３．学会等名
2022年度 人工知能学会全国大会（第36回）

４．発表年
2022年

１．発表者名
小田 直季，内田 智之，正代 隆義，松本 哲志，鈴木 祐介，宮原 哲浩

２．発表標題
順序木パターンの質問学習アルゴリズムによるグラフ畳み込みネットワークの予測根拠の可視化

３．学会等名
2022年度 人工知能学会全国大会（第36回）

４．発表年
2022年

１．発表者名
宮原 哲浩，鈴木 祐介，久保山 哲二，内田 智之

２．発表標題
ラベル情報を利用した進化的学習による複合的なワイルドカード付きタグ木パターンの獲得

３．学会等名
2022年度 人工知能学会全国大会（第36回）

４．発表年
2022年

１．発表者名
田中知希，鈴木祐介，内田智之，宮原哲浩

２．発表標題
頻出1変数項木パターンの枚挙アルゴリズム

３．学会等名
人工知能基本問題研究会SIGFPAI-120

４．発表年
2022年

１．発表者名
松本哲志，鈴木祐介，内田智之，正代隆義，宮原哲浩

２．発表標題
１つの正例と線形回数の所属性質問による変数次数が定数である線形順序項木パターンの言語族に対する質問学習アルゴリズム

３．学会等名
2022年電子情報通信学会総合大会

４．発表年
2022年

１．発表者名
門田大輝，鈴木祐介，内田智之，宮原哲浩

２．発表標題
物語文からの人物間の関係と動作を表す人物相関グラフ抽出手法の開発

３．学会等名
2021年度(第72回)電気・情報関連学会中国支部連合大会

４．発表年
2021年

１．発表者名
山本啓太，宮原哲浩，鈴木祐介，内田智之，久保山哲二

２．発表標題
進化的学習によるブロック内ワイルドカード付きブロック保存型外平面的グラフパターンの獲得

３．学会等名
2021年度(第72回)電気・情報関連学会中国支部連合大会

４．発表年
2021年

１．発表者名
Yuma Kawasaki, Tetsuhiro Miyahara, Tetsuji Kuboyama, Yusuke Suzuki and Tomoyuki Uchida

２．発表標題
Evolutionary Acquisition of Multiple TTSP Graph Patterns with Wildcards by Clustering TTSP Graphs

３．学会等名
2021 IEEE 12th International Workshop on Computational Intelligence and Applications（国際学会）

４．発表年
2021年

１．発表者名
徳原史也，沖永志帆，宮原哲浩，鈴木祐介，久保山哲二，内田智之

２．発表標題
ラベル情報を利用した進化的学習による複合的なワイルドカード付きブロック保存型外平面的グラフパターンの獲得

３．学会等名
2020年度 人工知能学会全国大会

４．発表年
2020年

１．発表者名
川﨑 有馬，宮原哲浩，山縣佑貴，徳原史也，鈴木祐介，内田智之，久保山哲二

２．発表標題
進化的学習による複合的なワイルドカード付きTTSPグラフパターンの獲得

３．学会等名
2020 IEEE SMC Hiroshima Chapter若手研究会

４．発表年
2020年

１．発表者名
酒井笑理，鈴木祐介，内田智之，宮原 哲浩

２．発表標題
1変数項木パターンに対するマッチングアルゴリズムの改良

３．学会等名
情報処理学会　第182回アルゴリズム研究発表会

４．発表年
2021年

１．発表者名
Fumiya Tokuhara, Shiho Okinaga, Tetsuhiro Miyahara, Yusuke Suzuki, Tetsuji Kuboyama, Tomoyuki Uchida

２．発表標題
Using Label Information in a Genetic Programming Based Method for Acquiring Block Preserving Outerplanar Graph Patterns with Wildcards

３．学会等名
2019 IEEE 11th International Workshop on Computational Intelligence and Applications（国際学会）

４．発表年
2019年

１．発表者名 松本哲志,正代隆義,内田智之,鈴木祐介,宮原哲浩
２．発表標題 Learning Algorithm for Erasing Regular Pattern Languages Using One Positive Example and a Linear Number of Membership Queries
３．学会等名 第174回アルゴリズム研究会
４．発表年 2019年

１．発表者名 舛井 里帆,池森 千尋,鈴木 祐介,内田 智之,宮原 哲浩
２．発表標題 1変数項木パターンに対する多項式時間マッチングアルゴリズム
３．学会等名 火の国情報シンポジウム2020
４．発表年 2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	内田　智之 (Uchida Tomoyuki) (70264934)	広島市立大学・情報科学研究科・准教授 (25403)	
研究分担者	正代　隆義 (Shoudai Takayoshi) (50226304)	福岡工業大学・情報工学部・教授 (37112)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関