

共同研究員の公募を行う共同研究について

1. 研究課題：基盤研究「歴史資料デジタルアーカイブデータを用いた知的構造の創生に関する研究—小袖屏風を対象として」

2. 研究代表者：横浜国立大学大学院工学研究院・教授 濱上 知樹

3. 研究期間：平成26年4月～平成28年3月（2年間）

4. 必要とする専門性と役割分担：

人文研究分野と知的情報処理の融合領域開拓に興味があり，下記の専門性のいずれかに該当する方（理系・文系を問わない）。研究成果を展示に結びつけていく上での実装技術支援やコーディネート，本融合領域を深化させる上で有益な知識・経験の共有をはかる役割を期待いたします。

- ・ 高精細画像処理および視覚心理学応用
- ・ セマンティックデータベース
- ・ 次世代デジタルアーカイブ技術
- ・ デジタル・キュレーション

5. 研究目的：

近年，急速に進む歴史資料のデジタルアーカイブ化により，貴重な歴史資料を計算機に取り込み，高度な知的利用や新たな展示技術として活用する研究が盛んに行われている。また，そのデジタルデータの取り扱いや記述・構造化について，メタデータの記述・モデル化の検討が進められている。このように，歴史資料のデジタルアーカイブの流れと関連研究は，人類共有の貴重な歴史資料の永続的な記録と再発見の手段として大きく期待されている。国立歴史民俗博物館においても，江戸図屏風を始め，正倉院文書，小袖屏風画等，数々貴重な歴史資料がデジタルデータ化され，資源共有する研究が進められてきた。

一方，近年の計算機技術とネットワークの飛躍的な発展を背景に **Big data** と呼ばれる超大規模データを取り扱う技術が注目されている。**Big data** では，従来のデータの取り扱いとは異なり，単に大規模というだけでなく，多様で非構造的なデータ発生源が時間的広がりを持って大量に存在する点に特徴がある。そのため，発生したデータを収集・蓄積する基盤技術と，収集・蓄積したデータを操作するための高度なツールが求められている。そのための重要な技術の1つが「**機械学習**」である。機械学習とは，データから有用な規則，知識表現，弁別基準，クラスタリングを抽出する人工知能の技術であり，蓄積されたデータの中から，有用な知識抽出を行う有効な手段として近年飛躍的な研究進展を遂げている。また，**Big data** 時代を切り開くもう1つの重要な技術が，1998年に提唱された**セマンティック Web** である。セマンティック Web とは「情報リソースに意味（セマンティック）を付与することで，人を介さずにコンピュータが自律的に処理できるようにするための技術」と定義される。メタデータとオントロジーによる自律的なデータの知的構造化，あるいは機械学習とセマンティックデータ技術の連携は，既に現実のネットワークの中で進み，様々な知的構造の構築とこれを用いた知的システムの構築に貢献をしている。

また、歴史資料にとって重要な要素であるイメージング処理技術についても、計算機処理能力の向上と共に目覚ましい高度化を遂げている。莫大な画像の中から意味のある情報や構造を抽出し、知識利用する技術は、医療、セキュリティ、アミューズメント等、社会の多くの場面で実用化されている。

翻って、増え続ける様々な歴史資料デジタルアーカイブー特に画像を中心としたデジタルデータとは、様々な非構造的特徴を有し、データ発生源の地理的・時間的拡がりを持つ Big data に他ならない。例えば、小袖屏風画像データベースに含まれる属性は、構図および衣装モチーフといった連続空間的属性から、染めや絞りといった技術的分類、さらには、時代や身分等、様々な外部属性につながっている。特に、研究上重要なモチーフ、染色方法、刺繍、文様のマイクロ・マクロな連続特徴量を扱うことから、非表象・非定型・非構造化された高精度画像情報からのセマンティクス抽出が重要な技術となる。しかし、このような**大規模な歴史資料に対し機械学習のような自律的データマイニングを試みた研究は過去に類がなく**、データ間のメタ構造を抽出する試みは緒についたばかりである。今後、**機械学習やセマンティックデータ技術をこれらの歴史資料に応用することで、新たな文理融合領域の研究が期待**できる。以上の前提から、本研究の意義・必要性を以下のように定める。

**研究の意義** 歴史資料デジタルアーカイブデータに最新のイメージング処理をフロントエンドとする機械学習・セマンティックデータ処理を試みることで、デジタルアーカイブデータからの新たな知的構造の創出と、データマネジメント技術の発展をはかり、新たな学際領域の発展と歴史資料を用いた視覚的情報提示手法の確立ー**歴史資料高度データマネジメントシステム**（仮称）の構築を行う。

**研究の必要性** 歴史資料デジタルアーカイブの蓄積が進む中、その莫大なデータの利活用の方法が重要となっている。電子的に保存されたデータベースとしての静的な存在ではなく、非構造化されたデータから意味や新たな構造を抽出する自律性を付与することで、歴史資料の新たな利活用とこれを用いた研究の発展が期待できる。本研究では、**歴史資料デジタルアーカイブを Big data とみなし、様々な分野で活用されている画像処理・解析を初めとする特徴抽出と、これらの解析技術で培われた機械学習、セマンティックデータ処理を用いて、莫大なデジタルアーカイブからの知識発見をはかるシステムをつくる**ことで、この期待に応えるものである。

この研究の嚆矢として、まず**野村コレクション小袖屏風のデジタルアーカイブデータについて検討を開始**する。小袖屏風画像は、既にデジタル化がなされており、画像の様々なレベルにおける特徴量の抽出とその構造化を進め、さらに他の史資料についても検討を拡大する。

以上の試みにより、**歴史資料デジタルアーカイブデータを用いた知的構造の創生をはかり、歴史研究の新たな展開と、得られた成果の効果的な展示技術に結びつける**ことを目的と定める。

## 6. 研究計画：

平成25年度（済）

初年度は、小袖屏風画像の高精度デジタルアーカイブデータからのメタ構造の抽出準備と試行を重ね、次年度以降に着手する、**歴史資料高度データマネジメントシステム（仮称）**の設計指針をつくることをめざす。具体的には、下記の事項を明らかにする。

1. 画像メタ情報の自律的抽出アルゴリズムの検討（濱上ら知能化班）
2. 高精度画像データからの特徴抽出アルゴリズムの検討（内田・中島ら画像処理班）
3. 高精度画像セマンティックデータベースの技術要件に関する検討（富井らデータベース班）
4. 次世代デジタルアーカイブのための要件整理（福永らアーカイブ展示班）

それぞれ個別に検討を進めながら、1と2についてはあわせて小袖屏風画像データを用いた実分析を試行し、その結果をふまえて3、4の以上の展望をさらに深く検討する準備とする。これらの成果を研究会・各関連学会、研究会等にて公表し、諸関係者からの意見を収集する。

#### 平成26年度

前述の個別検討の成果をうけて、**歴史資料高度データマネジメントシステム（仮称）**の設計に着手する。本システムを、従来のメタデータ記述データベースのスーパーセットとして位置づけ、Big data 分析、知能化処理、高度画像処理エンジンを行うための分散環境として設計する。このシステムは、医療データにおけるDICOMやLevel7のように、あるドメインの広範なデータ構造を内包する仕様をもち、様々な史料研究に資するサブシステムが連携することで高度な画像処理、知能化処理、データ処理、ネットワーク共有化をはかることをめざす。このように様々なデータが分散、連携するしくみをつくることによって、これまで蓄積された膨大な史資料データの利活用と、今後発生するさらに**大規模高精度な資料管理を行うフレームワーク**とする。この段階でも、小袖屏風画像を中心としたプロトタイプシステムの完成をめざすが、25年度検討で進められる次世代デジタルアーカイブのための要件定義にもとづいて、新たな史資料の収集と魅力的な展示方法に資する技術的課題についても、個別に検討をすすめる。

#### 平成27年度

歴史資料高度マネジメントシステム（仮称）プロトタイプに基づき、ここまでの段階で明らかになった効果や問題点を整理し、本システム設計のための提言をまとめる。また、得られた成果を効果的な展示、研究者支援に活用していくための具体的なアプリケーションのプロトタイプ実現をはかる。

以上の各年度成果を、歴博における研究会、関連学会にて逐次公表し、**人文学研究者の意見も広くとりいれながら、歴史民俗博物館ならではの人文・理工融合分野の開拓と、今後の歴史資料デジタルアーカイブの利活用をはかるオープンプラットフォームとして完成**を計画している。

7. 研究成果の公開方法：「研究報告」及び単行本等の刊行、各種学会におけるシンポジウム、企画セッション等