位置情報とトピックモデルに基づくフィールドノートのビジュアライズ手法

高田 百合奈
首都大学東京大学院
システムデザイン研究科

渡邉 英徳
首都大学東京大学院
システムデザイン研究科

柳澤雅之
京都大学地域研究
統合情報センター

山田太造
東京大学史料編纂所
前近代日本史情報国際センター

研究者によってフィールド調査により記録されたデータは，アーカイブされフィールドノートとしてまとめられている．しかしこれまではこれらの文章群が記録された紙媒体を閲覧することができるのみで，そこから文章の分析をすることでしかフィールド調査の結果を導き出すことが出来なかった．したがって本研究は，フィールドノートから，その土地の情報が一見して理解できる形に可視化することを目的とし，研究を行った．
フィールドノートの記録を取った場所を全体文や前後関係から推測し，Google Earth 上のその場所にデータをマッピングする．さらにデータは，データマイニングにより，各々のデータを色別で表現することで，記録した場所とその土地の情報が可視化できるシステムを開発する．Google Earth API を用いて，web コンテンツとして実装する．本システムにより，フィールドノートによる記録を位置情報とトピックモデルに基づいて可視化することが出来る．

A Visualization Method of Field Notes based on Locations and Topic Models

TAKATA Yurina
Graduate School of
System Design, Tokyo
Metropolitan University

WATANAVE Hidenori
Graduate School of
System Design, Tokyo
Metropolitan University

YANAGISAWA Masayuki
Center for Integrated Area
Studies, Kyoto University

YAMADA Taizo
Historiographical Institute,
University of Tokyo

The data recorded by researchers at field research are archived and gathered up as field notes. It is just possible to view these data in writing until now, we can derive a result of field research only by qualitative analysis of the data. Therefore, the purpose of this study is to visualize the local features data in a comprehensive form from field notes data.
We guess a recording position of a data from whole sentence and the data is mapped at the point on Google Earth. We develop a system to visualize recording positions and land information with displaying the data in each color by data mining. This system is implemented as web contents using Google Earth API. We can visualize the field notes data based on locations and topic models by this system.

Keywords / field notes, topic model, visualize, mapping, Google Earth

(c) Information Processing Society of Japan

1．本研究の目的

研究者がフィールドワークの中で記録した，観察や聞き取り調査のメモは，フィールドノートとしてまとめられている[1][2]．これらの記録は，テキストデータ化され，そこから分析，及び考察を行った研究は多数見られる[3]．このような研究では，これまではフィールドノートをテキストデータ化してアーカイブし，テキスト情報として閲覧出来るのみで，そこから文章の分析をすることでしか，フィールド調査の結果を導き出すことが出来なかった．そのため，そこから地域の特徴を把握することは，専門家でなければ困難であり，さらに他の地域との関係性を理解することには適していないと考察する．

フィールドノートをデジタルアーカイブする試みとして，iPad を利用し，手書きの記録を書き込むことでデジタル地図にマッピングする，フィールドノートシステムを開発した事例[4]がある．これより，フィールドノートはデジタル地図へのアーカイブに適していると考察できる．しかし，過去にアーカイブされたフィールドノートのデータをマッピングするための方法は開発されておらず，またマッピングされたデータを 1 つずつ確認しなければ，データの中身を把握することができないという問題がある．

したがって本研究は，フィールドノートから，その土地の情報が一見して理解できる形に可視化することを目的とし，ビジュアライズ手法の提案を行う．

2．関連研究

フィールドノートのデータは，文章内に地名や距離情報，時間情報を含んだ状態で，時系列上に陳述されている．位置情報を含んでいることより，適したビジュアライズ手法について既存研究を参照しながら検討する．

高田らが開発した，海洋生態系の情報を Google Earth 上にマッピングすることで，海洋生態系の経年変化の可視化を試みた「プロジェクト・ヨロン」がある[5]．珊瑚やなどの海洋情報をそれぞれ分類し，種類ごとにアイコンを設け，Google Earth 上にマッピングしている．地形と重層表示させることで，地形との相互参照や，マッピングされている他のデータとの比較検討から，データ同士の関連性についても考察する事ができる．他にも渡邉らの「Nagasaki Archive」[6]や戸根らによる「映像・画像資料アーカイブ連携・時空間処理システム」の研究[7]など，多数の種類のデータをデジタル地図上に重層表示した事例があり，地域情報を地図上にアーカイブする事は，情報の理解や，データ同士の関連性の理解に有効であると考える．従って本研究では，フィールドノートのデータをデジタル地図上にアーカイブする事とする．

しかし，データを種類別にマッピングするだけでは，アイコンをクリックしなければその場所の情報が見えないため，一見して地域の情報を把握することは困難である．そこで，デジタル地図上での迅速なデータの理解を可能とするため，フィールドノートの文章群をデータマイニングし，分析結果を元にマッピングする手法について検討する．

ここで，データマイニング手法の 1 つである，トピックモデルを参照する．トピックモデルとは，文章全体の中で起こっている，トピックと呼ばれる概要を見つけるための手法である．文章内に含まれる，特徴語句の出現確率の組み合わせにより，他の特徴語句との関連の度合いを導き出すことができる．トピックモデルを利用することで，ある 1 つの文章が何について書かれた文章なのかクラスタリングすることや，特徴語句自体をどのような潜在的な意味を持っているのかクラスタリングする事ができる．これによって，文章の内容や，文章内での単語の使われ方などの理解が可能となる．

したがって本研究では，データの種類別にアイコンを設けてデジタル地図上にマッピングし，さらにトピックモデルの手法を用いて，フィールドノートのデータをデータマイニングし、デジタル地図上にアーカイブすることで，フィールドノートのビジュアライズ手法を提案することとする．また本手法によって，「フィールドノート・アーカイブ」コンテンツとしてシステムを実装する．本稿では本システムの実装例を通して，本手法について述べる．

3．提案する手法

本研究では実装例として，京大の故高谷好一氏のフィールドノートを利用する．高谷氏は長年のフィールドワークで東南アジア全域を調査しており，そのデータはテキスト化されている．テキストデータ化した冊子は何冊にも及んでおり，本研究ではこの中の 1 冊である，インドネシアのフィールドノートを実装することとする．フィールドノートの中身は，景観写真やスケッチ，聞き取り調査のメモ等多様にあり，文章内に地名や距離情報を含んだ状態で，時系列上に陳述されている．これらの情報をデジタル地図上にビジュアライズすることで，その土地の理解が容易になると考え，次の工程に沿って実装する．

(c) Information Processing Society of Japan

第一に，どこで記録した情報なのか昔の地図を見ながら推測し，手作業で位置情報を付加する．このデータを csv データにまとめ，Google Earth で読み込めるデータ形式である kml に出力する．Google Earth API を用いて kml を読み込み，javascript で web コンテンツとして実装する（図1）．

次章より，具体的な実装例について述べる．

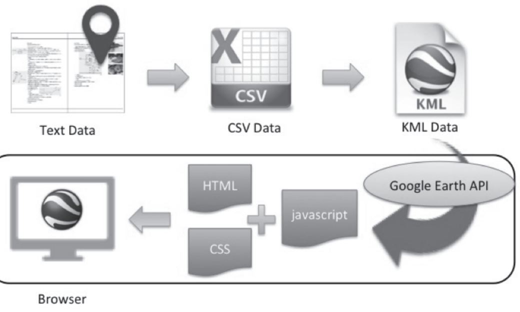

図 1　システム構成図
Figure1　System Configuration

４．アイコンによるマッピング

風景，土地利用，ヒアリング，場所，まとめ，その他の 6 種類に分類し，それぞれにアイコンを設定する．マッピングされているアイコンをクリックすると，そのポイントの記録である，文章や写真が表示される．また，データは記録が取られた順番に ID を振り，時間データも付加させることで，その土地に関するデータの前後関係も可視化できる（図 2）．

しかしこれだけでは，アイコンをクリックしなければその場所の情報が見えないため，地域の情報を把握することは困難である．そこで，フィールドノートの文章群に，トピックモデルを利用したデータマイニングを施し，カテゴライズされたデータを地図上に可視化することで，地域情報の把握を可能とするシステムを開発する．

５．トピックモデルによるビジュアライズ

本研究では，フィールドノートのトピックモデルを，東京大学史料編纂所が作成し，それを応用することとする．また，今回利用するトピックモデルのトピック数は 30 種類である．

５．１ 色別トピックモデルの実装

30 種類のトピックは，トピック 1 からトピック 30 に分類されており，記録の文章内に出現する特徴語句と文章自体が，30 種類のうちどれか

図 2　アイコンによってマッピングしたコンテンツ画面
Figure2　Content of System mapping data by icons

(c) Information Processing Society of Japan

にカテゴライズされている．この 30 種類のトピックを色別で示すことにより，色から文章の持つテーマを理解することが出来ると考察する(図3)．つまり，色別のトピック表示は，その土地の特徴を理解することに役立つと考える．そこで，トピック同士の関連性が高いほど，近い色相になるよう設定し，色別トピックモデルを生成する．この色別トピックモデルを Google Earth コンテンツに応用させることで，俯瞰的な土地情報の理解を可能にすると考察し，検索システムとして実装することとする．

次節より色別トピックモデルの具体的な実装例について述べる．

５．２　色別トピックモデルによる検索システム

トピックの一覧から単語を選択，もしくは文章中の単語を選択することで，同単語が含まれるポイントを抽出し，地図上に同心円画像を重層表示する．以下に挙げる 2 種類の検索機能を実装する．

（1）トピックの種類から検索

図 4 はトピックから検索した結果の画面である．右の画面のカラーチャートから 1 つトピックを選択すると，そのトピックにクラスタリングされている文章が出現するポイントを抽出する．このとき，選択したトピックの色の同心円画像が，そのポイントを中心にオーバーレイして表示される．これによって同じような特徴を持つ土地がどこに広がっているのかを理解することが出来る．

（2）特徴語句から検索

特徴語句の一覧から 1 つ特徴語句を選択すると，選択した語句と同じ語句（トピックも同じ）が出現するポイントを抽出する．同じ単語であっても，異なるトピックを持つ文章内で出現することもあるため，このような機能を設ける．この時，それぞれのポイントにおける文章のトピックの色の同心円を表示する．これによって同じ単語が，どのような土地で出現するのかを可視化することが出来る．

以上より実装した，トピックモデルを利用したフィールドノートのビジュアライズシステムの効果を計るため，次章では本システムを実際に利用し，そこから得られる知見について述べる．

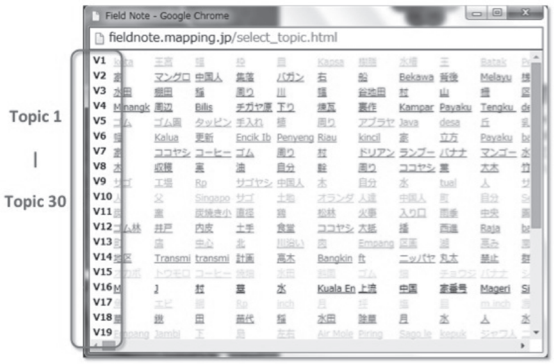

図 3　30 種類のトピック一覧
Figure3　Table of the topics

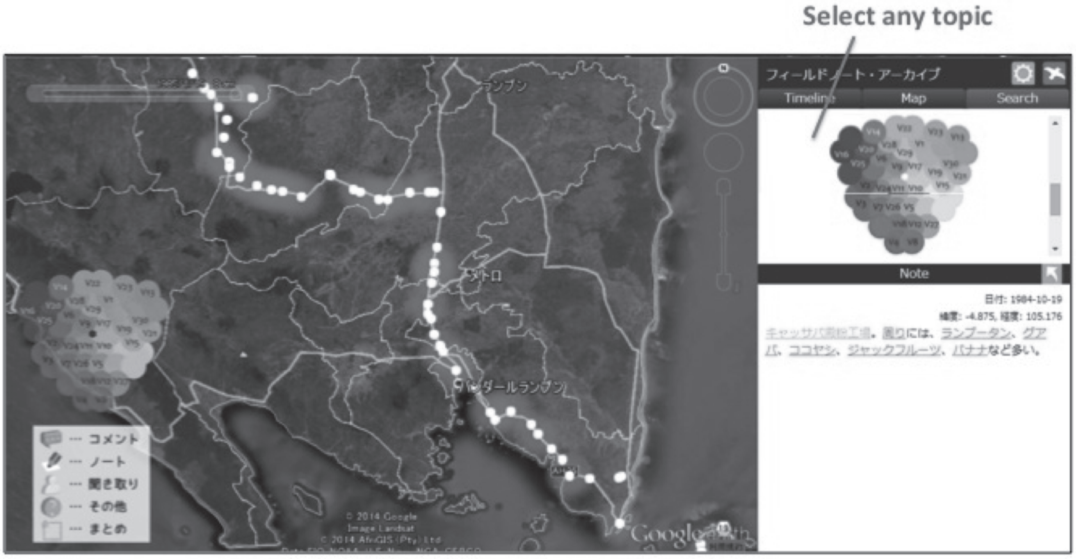

図 4　トピックの種類から検索した結果
Figure4　The result of searching from Topics

(c) Information Processing Society of Japan

６．検証

６．１　トピックの種類からの検索

図 5 は, オレンジ色のトピック 5 を検索した結果である．このように，地域によって偏りが見られ, トピック 5 の傾向にあるところとないところが判明する．このように，Google Earth の 3D の地形上に重層表示されることで，川や山，市街地といった地図上から理解できる土地情報と，フィールドノートの記録から読み取った土地情報との関係性を探ることに貢献できると考える．

６．２　特徴語句からの検索

次に，ピンク色のトピック 7 の「コーヒー」という単語を検索した結果を示す（図 6）．トピック 7 の「コーヒー」が含まれた文章が出現する場所が抽出される．この場合は，図 7 で表示されている地域は全てのピンク色で示されているため，トピック 7 の「コーヒー」は，同じ特徴を持つ土地に出現していることが分かる．

一方図 7 は, 緑色のトピック 15 の「コーヒー」を検索した結果である．図 6 とは異なり，多くの色で示されている．これは，同じトピック 15 の「コーヒー」でも，様々な特徴を持つ土地に出現していることを意味する．つまり，トピック 7 の「コーヒー」は同じような特徴を持つ場所に出現するが，トピック 15 の「コーヒー」は，様々な土地と関係があるということを示していると考察する．

さらに，同じトピック 15 の「トウモロコシ」を検索する（図 8）．この場合も同じトピック 15 の「トウモロコシ」でも，様々な特徴を持つ土地に出現することが分かる．つまり，トピック 15 自体が，他の特徴を持つ土地と関係があり，単独では存在しないということを意味していると考察できる．

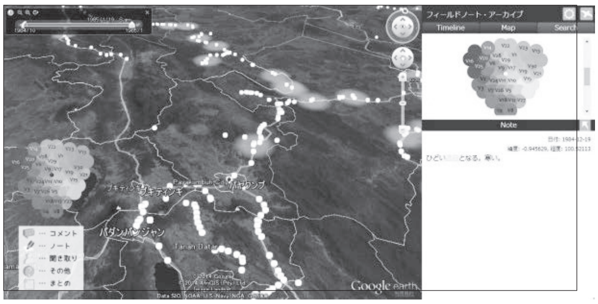

図 5　トピック 5 の検索結果
Figure5　The searching result of Topic 5

図 6　トピック 7 のコーヒーの検索結果
Figure6　The searching result of “coffee” in Topic 7

図 7　トピック 15 のコーヒーの検索結果
Figure7　The searching result of “coffee” in Topic 15

図 8　トピック 15 のトウモロコシの検索結果
Figure8　The searching result of “corn” in Topic 15

(c) Information Processing Society of Japan

7．まとめ

本研究では，フィールドノートによる記録を位置情報とトピックモデルに基づいて，Google Earth 上に実装した web コンテンツを制作した．本システムの検索機能を試したところ，俯瞰的に地域の特徴を把握でき，さらに他地域との関連を考察することが出来た．他のトピック及び特徴語句で検索すると，農業の仕方や，人の活動を発見することが出来ると期待しており，今後本手法の効果について実証実験を行う方針である．

また本手法は，フィールドノート以外のデータの可視化にも活用でき，その土地の経年変化の観察や，ユーザ行動の可視化等にも応用できると考えている．

参考文献

1) 須藤護：フィールドノート：東北アジアにおける「熊民俗」の予備的研究；龍谷大学国際社会文化研究所紀要，Vol.15，pp.81-109（2013）．

2) 大島圭子：モロッコ都市のフィールド・ノートから；大東アジア学論集，Vol.1，pp.73-75(2001).

3) 松島茂：華南における産業集積の変容の可能性：玩具関連企業のフィールドノートからの考察;経営志林, Vol.41, No.4, pp.137-145(2005).

4) 菊池佑太，加藤直樹，山崎謙介：野外学習のための Web-based フィールドノートシステムの開発；情報処理学会研究報告，Vol.130，No.75-82（2006）．

5) 高田百合奈，朴婉寧，蜂谷聖未，高田健太郎，西田志帆，渡邉英徳，植田佳樹：デジタル地球儀を用いた成長型海洋生態系アーカイブのデザイン手法；日本バーチャルリアリティ学会論文誌，Vol.17， No.3，pp.253-260（2012）．

6) 渡邉英徳，坂田晃一，北原和也，鳥巣智行，大瀬良亮，阿久津由美，中丸由貴，草野史興：“Nagasaki Archive”：事象の多面的・総合的な理解を促す多元的デジタルアーカイブズ；日本バーチャルリアリティ学会論文誌，Vol.16， No.3，pp.497-505（2011）．

7) 戸根嘉元，岡田至弘：映像・画像資料アーカイブ連携・時空間処理システム；情報処理学会研究報告．人文科学とコンピュータ研究会報告，Vol.2013， No.1，pp.1-6（2013）．

(c) Information Processing Society of Japan