前近代日本史史料における人物関係とその時空間変化：天正期古記録『上井覚兼日記』を例に

山田 太造（東京大学 史料編纂所／地震火山史料連携研究機構）
畑山 周平・小瀬 玄士・遠藤 珠紀・井上 聡・久留島 典子
（東京大学 史料編纂所）

本論文では，前近代日本史史料に登場する人物間の関係を史料での共起関係をベースとして，時間・空間の観点から分析していく手法を導入する．共起関係を検出するため，トピックモデルの 1 つである LDA を用いた．天正期古記録『上井覚兼日記』を対象に提案手法の検証を行った．その結果についても述べる．

Personal Relations in Per-modern Japanese Historical Materials and their Spatiotemporal Changes: A Case Study of the Old Diary in Tensho Period "Uwai Kakuken's Diary"

Taizo Yamada
(Historiographical Institute / Collaborative Research Organization for Historical Materials on Earthquakes and Volcanoes, The University of Tokyo)
Shuhei Hatayama / Genshi Kose / Tamaki Endo / Satoshi Inoue / Noriko Kurushima
(Historiographical Institute, The University of Tokyo)

In the paper we introduce a method which can analysis relationships between persons based on the co-occurrences in pre-modern Japanese materials and the spatio-temporal variations. In order to detect the co-occurrence relationships, we used LDA which is one of topic models. We verified our method with the old diary in the Tensho period "Uwai Kakuken's Diary" as the case study. Also we show the results.

１．はじめに

歴史的事象を直接観察することは困難である．歴史的事象の理解やその構成を行うため，観測された事実から理解の対象である歴史的事象を推定していくことは可能である．これは日本史を含む歴史学の研究の過程に相当する．観測された事実の 1 つとして史料が挙げられ，史料収集，読解，史料批判は歴史学研究の基礎として位置づけられる[1]．我々はこれまで[2,3,4]などの取り組みを介して，歴史的事象の推定・理解を推進していくための切り口として，収集した史料から人物・空間・時間などのデータを生成・管理してきた．本研究では，このデータ構成のバリエーションにより，この切り口を洗練することを目的とする．史料は，所在・伝来，伝来や管理などによる形態，史料群の形成，性格，素材，様式，文字，内容，人物，空間，時間などの多様なパラメータにより特徴づけられる．人物間の関係を見出すため，単純に，史料に関係する人物・地名，および史料の年代を扱うことにする．これにより人物関係の時系的／空間的変動に伴う変化を知りうる方法を導入する．

２．人物関係

史料に記載された内容を読解することによりそこに登場する人物・地名・時間を知ることができる．本研究では，これらのパラメータにより，史料の話題を特徴づけることにした．史料には複数の人名が記載されていることが多々ある．そこ

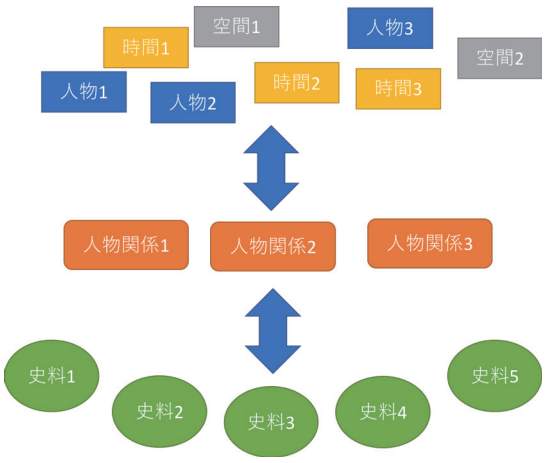

図 1　史料と関連データ
Figure 1　Historical materials and related data

©2017 Information Processing Society of Japan

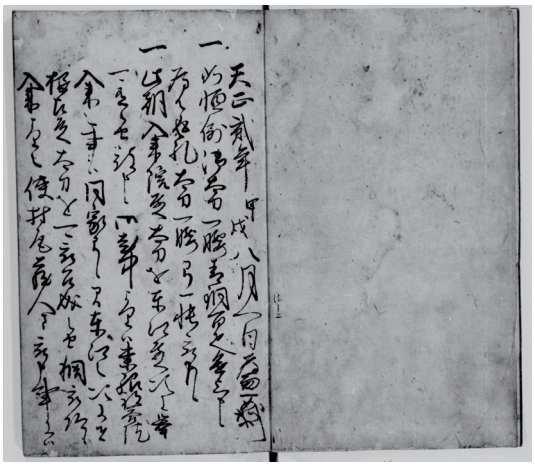

図 2　『上井覚兼日記』天正 2 年 8 月 1 日条
Figure 2　A record of Aug 1 Tensho 2 in “Uwai Kakuken’s Diary”

で，史料本文に出現する人名を単に扱うのではなく，人名の共起関係を用いることで，潜在する意味関係を検出する．これにより人物間の関係性を見出し，さらに，出現する地名および時間を加えることで人物関係を表現する（図 1）．ここで，史料および人名・地名・時間は観測できるが，人物間の関係性は明示されていない．この検出を行うため，トピックモデルの 1 つである LDA（Latent Dirichlet Allocation）[5]を用いた．また，本研究では天正期古記録『上井覚兼日記』を対象に，この効果の検証を進めた．

２．１．『上井覚兼日記』

『上井覚兼日記』（図 2）は薩摩島津家の家臣であった上井覚兼による日記である．原本は東京大学史料編纂所所蔵の重要文化財であり，国宝『島津家文書』に含まれる史料である．期間は天正 2 年 8 月 1 日から天正 14 年 10 月 15 日までの 12 年に及ぶ．書かれた期間および著者である上井覚兼の立場から，天正 2 年 8 月 1 日から天正 4 年 9 月 6 日までの鹿児島における奏者時代，および天正 10 年 11 月 4 日から天正 14 年 10 月 14 日までの日向国宮崎城主時代の 2 つに分けることができる．『上井覚兼日記』は『大日本古記録 上井覚兼日記』（全 3 冊）として東京大学史料編纂所（以下，史料編纂所）より編纂史料集が発刊されており，その全文は史料編纂所データベース検索サービス SHIPSDB[1]の 1 つである「古記録フルテキストデータベース」（以下，古記録 FT）に，この全テキストが格納されている．『上井覚兼日記』は古記録 FT では日条を単位として格納してあることから，これを分析の単位とした．日条は 1777 件である．

[1] http://wwwap.hi.u-tokyo.ac.jp/ships/

図 3　人名リポジトリ
Figure 3　Personal name repository

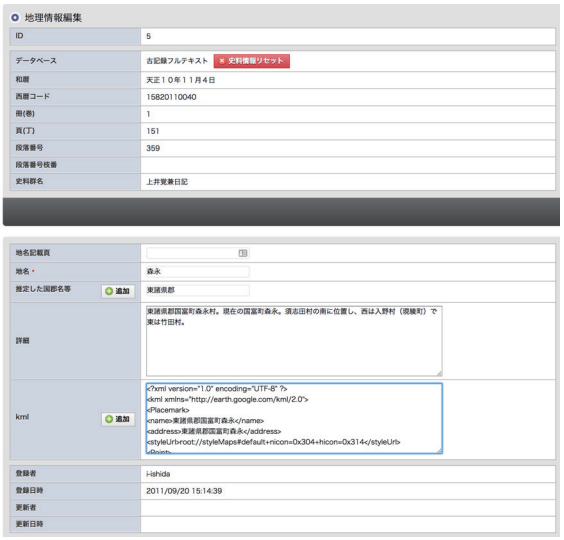

図 4　地名リポジトリ（地名基盤管理システム）
Figure 4　Place name repository (Management system of place name)

２．２．人名データ

我々はこれまでに，史料に登場する人物の情報を格納するために人名リポジトリ[2]を構築してきた．人名リポジトリでは，ある人物に対して，よく知られている人名とその人名が登場もしくは関係する史料データ，および史料上の表記を対応づけたデータを格納することができる（図 3）．人名データに関して，JapanKnowledge[8]・コトバンク[9]などのオンラインサービスで提供されている「日本人名大辞典」や Wikipedia における人物の記事のような人名辞典がある．一方で国立歴史民俗博物館「地域蘭学者門人帳人名[10]」・「江戸商人・職人[11]」，国文学研究資料館「歴史人物画像データベース[12]」などがある．これらは，それぞれのデータベースで対象としている史料から人名を網羅的に収集し，人名とそれに関連す

©2017 Information Processing Society of Japan

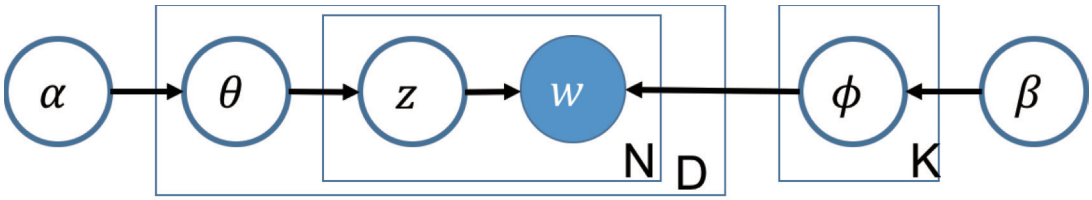

図 5　LDA のグラフィカルモデル
Figure 5　A graphical model of LDA

る史料の典拠などが提示されている．人名リポジトリは，人名辞典を作成していくのではなく，人名をその典拠となる史料とともに，時代や地域を横断的に，かつ網羅的に収集し，管理していくことを主眼としている．

『大日本古記録』ではその編纂過程において，史料に登場する人物に対してよく知られている人名を付与しており，古記録 FT においてもそれが表現されている．この人名データを人名リポジトリに格納していくことで人名−史料関係を格納している．また，人名リポジトリでは格納されたデータを RDF（Resource Description Framework）形式で管理・共有することができ，また SPARQL endpoint としての機能を有している．これにより，システムに依存せずにデータを検索・取得することが可能である．

２．３．地名データ

2.2 節と同様に，我々は，史料に出現した地名を蓄積する地名リポジトリを構築してきた．地名データとして，JapanKnowledge 等で提供されている「日本歴史地名体系」，国立歴史民俗博物館「日本荘園[13]」「旧高旧領取調帳[14]」などがある．地名リポジトリは，人名リポジトリと同様に，地名辞典を作成するのではなく，時代・地域を横断し，史料に出現する地名を網羅的に収集することを目的としている．

地名リポジトリでは表記上の地名，推定した国郡名，緯度経度，史料データ等を蓄積することができる（図 4）．人名リポジトリと同様に，RDF ストアかつ SPARQL endpoint として機能する．一部であるが『上井覚兼日記』における地名についても，格納している（天正 10 年 11 月から天正 12 年 4 月まで）．本研究では格納されているこのデータを利用した．

２．４．人物間の関係性

人物間の関係性を検出するため，トピックモデルの 1 つである LDA を用いた．まず史料ごとに出現する人名を BOW (bag-of-words) で表現する．このとき史料からの人名の取得では人名リポジトリを用いた．人名の BOW に対して LDA を適用する．LDA は予め設定したトピック数に応じて，人名を分類していく．

1. Initialize α and β
2. Initialize z
3. Set S : the number of sampling
4. for s = 1, …, S do
5. for d = 1, … , D do
6. for i = 1, … , N_d do
7. Sample $z_{d,i}$
8. Update $N_{d,z_{d,i}}$
9. end for
10. end for
11. Update α and β
12. end for

図 6　崩壊型ギブスサンプリングによる推定手続き
Figure 6　Procedure of inference by Collapsed Gibbs Sampling

本研究で用いた LDA のグラフィカルモデルを図 5 に示す．ここで青色の円は観測変数，白色の円は未知変数を示す．矩形は繰り返しを，矩形にある右下の数値は繰り返しの回数を示す．wは唯一の観測変数である人名を示す．zはwのトピックを示す．θは日条ごとのトピックの確率分布，ϕは全人名でのトピックの確率分布を示す．αおよびβはそれぞれ，θおよびϕのパラメータを示し，このモデルではハイパーパラメータに相当する．日条数をD，日条dの人名数をN_d，トピック数をKとしたとき，θ_dおよびϕ_kは

$$\theta_d \sim Dir(\alpha)\ (d = 1, \dots, D),$$
$$\phi_k \sim Dir(\beta)\ (k = 1, \dots, K) \qquad (1)$$

により生成される．ここで，$Dir(\cdot)$はディリクレ分布を示す．トピック$z_{d,i}$は下記により生成される．

$$z_{d,i} \sim Multi(\theta_d) \quad (i = 1, \dots, N_d) \qquad (2)$$

ここで$Multi(\cdot)$は多項分布を示す．さらに人名$w_{d,i}$は

$$w_{d,i} \sim Multi\left(\phi_{z_{d,i}}\right) \quad (i = 1, \dots, N_d) \qquad (3)$$

により生成される．

LDA のモデル推定では崩壊型ギブスサンプリング(Collapsed Gibbs Sampling; CGS)による解法が知られている[6]．この手続きを図 6 に示す．ここ

©2017 Information Processing Society of Japan

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼
2	本田親貞	島津義久	吉利忠澄	柏原有閑	島津歳久	吉利忠澄	比志島義基	島津義久	島津義弘	島津家久
3	平田増宗	伊集院忠棟	秋月種実	鎌田兼政	本田薫親	上井薫兼	伊集院久宣	伊集院久治	上原尚近	島津義弘
4	白浜重治	伊地知重秀	龍造寺政家	上井兼成	阿多忠辰	肝付兼寛	吉利久金	島津忠長	島津忠長	伊集院忠棟
5	伊地知重秀	北郷時久	高城珠長	野村豊綱	喜入季久	鎌日兼政	平田宗張	川上久隅	島津家久	三田井親武
6	八木昌信	本田親貞	大友義統	上井薫兼	土持久綱	伊集院忠棟	新納久時	島津義弘	島津義久	上井神九郎
7	肝付兼寛	喜入季久	島津歳久	桑幡道隆	敷祢頼賀	有馬晴信	大寺大炊助	伊集院忠棟	山忠助	上井秀秋
8	伊集院忠棟	島津以久	稲長辰	田尻鑑種	島津以久	大寺大炊助	川上翌久	山有信	名和顕孝	柏原有閑
9	島津忠長	平田昌宗	有馬晴信	平田宗張	川上久辰	島津義久	山忠助	本田正親	平田歳宗	有川貞真
10	平田光宗	島津義弘	筑紫広門	敷禰頼元	平田光宗	平日光宗	吉利忠澄	本田親正	新納久時	本田正親
11	本田薫親	島津朝久	上原尚近	土持久綱	赤星統家	島津忠長	山有信	合志親重	伊集院久治	税篤和
12	吉田清存	伊地知重隆	井尻祐貞	平田宗応	宮原景種	平日宗衡	上井秀秋	村田経平	伊集院忠棟	島津忠長
13	長谷場純辰	平田宗張	島津義弘	新納忠真	島津家久	敷祢頼賀	吉田清存	猿渡信光	大野忠宗	志岐鎮経
14	税篤和	村田経定	平田宗位	三池鎮実	有馬晴信	鎌日政景	吉利忠富	喜入季久	新納教久	新納忠包
15	龍造寺隆信	東郷重尚	伊地知重則	三田井親武	稲長辰	奈良原延	大寺政安	島津忠良	三原重益	本田親商
16	上原尚近	河野通貞	鎌田政虎	市来忠末	本田盛親	島津家久	小川有季	新納久饒	北郷忠虎	野村友綱
17	木脇祐充	国分定友	織田信長	桂忠詮	本田親兼	肝付兼篤	市来家守	川田義朗	宮内忠増	島津久保
18	羽柴秀吉	肝付兼寛	近衛前久	平田宗清	猿渡信光	伊地知重元	柏原有閑	小代親泰	山有信	伊集久治
19	三原重隆	伊地知重興	伊集院忠棟	猿渡信光	鎌田政景	大山綱秀	比志島国貞	川上忠智	島津歳久	北郷久盛
20	島津貴久	本田親治	志岐親重	猿渡信孝	高城珠長	川上忠智	稲長辰	平田光宗	本田正親	吉利久金

図 7　LDA よるトピック検出の結果
Figure 7　Result of topic detection by LDA

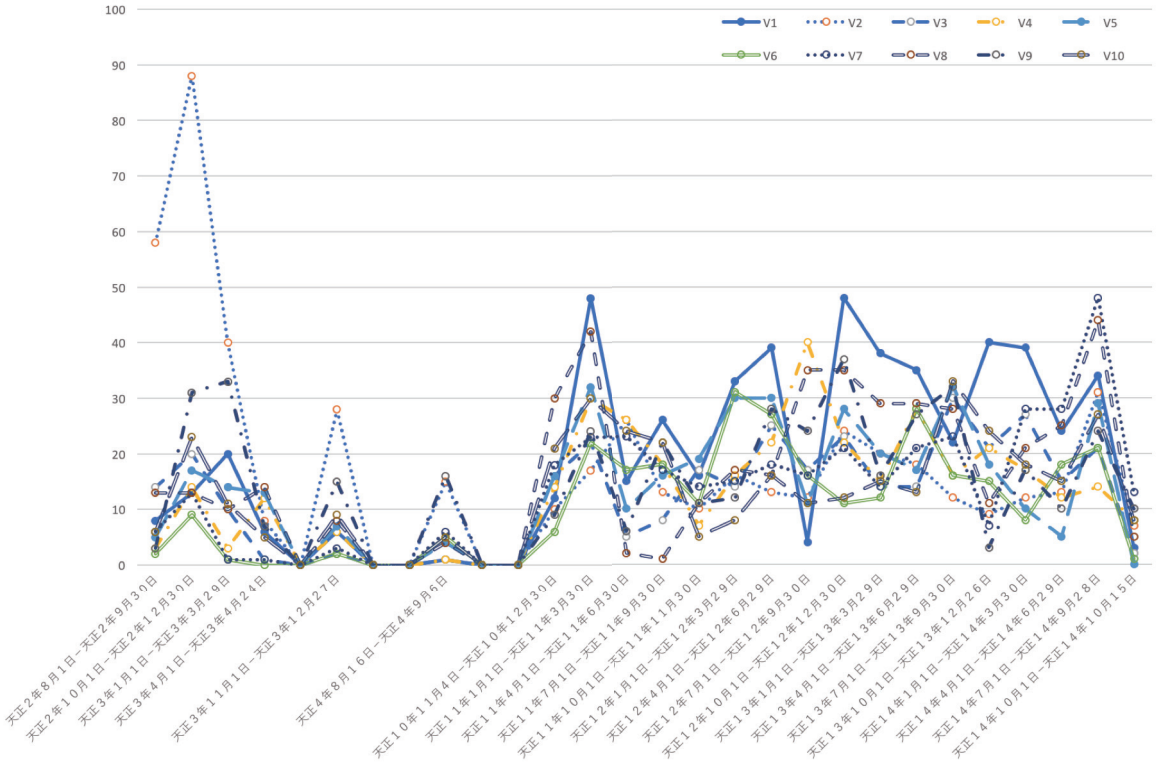

図 8　トピックの時系列変化
Figure 8　Time series changing of topics

で$N_{d,i}$は日条dにおいてトピックiに割り当てられた人名数，Vは人名数を示す．CGS ではθおよびϕを周辺化（積分消去）することにより直接求めることなく，次式によりzをサンプリングすることができる．

$$z_{d,i} \sim Multi\left(p\left(z_{d,i}|W, Z_{\backslash d,i}\right)\right)$$
$$\propto \left(N_{d,i} + \alpha\right)\frac{N_{d,iw_{d,i}}+\beta}{N_k+\beta V} \qquad (4)$$

LDA のハイパーパラメータであるαおよびβはそれぞれθ，ϕのパラメータである．[7]により，α

©2017 Information Processing Society of Japan

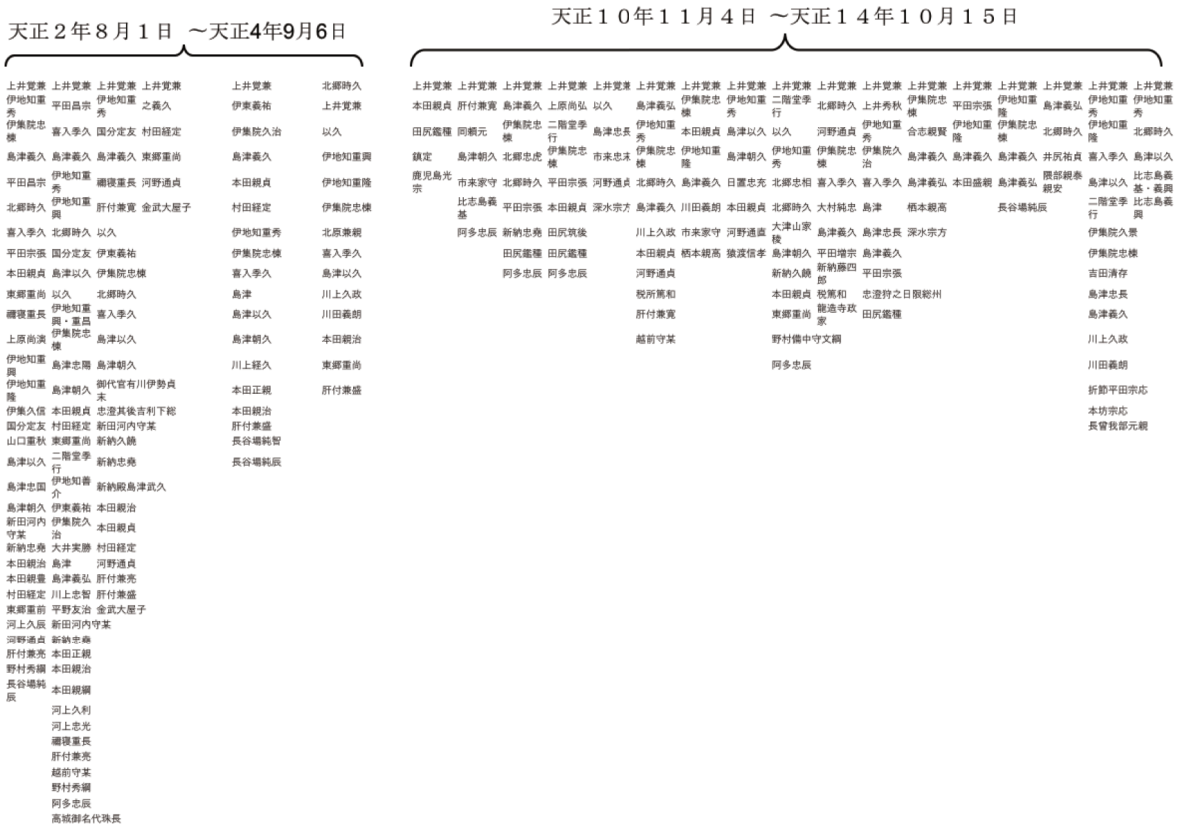

図 9　トピック 2 の時系列変化
Figure 9　Time series changing of Topic 2

は一様ではなく（$\alpha_k \neq \alpha_l, k \neq l$），$\beta$は一様である（$\beta_1 = \beta_2 = \cdots = \beta V$）である場合，LDA のパフォーマンスが改善されることが報告されているため，本研究でも同様の設定とした．さらに CGS の各サンプリングにおいてαおよびβを更新する．次式に示すように，不動点反復法によりディリクレ分布におけるパラメータを推定することができる．

$$\alpha_k^{new} = \alpha_k \frac{\sum_d \Psi(\mathrm{N_{d,k}}+\alpha_\mathrm{k}) - \mathrm{D}\Psi(\alpha_\mathrm{k})}{\sum_d \Psi(N_d + \sum_{k'} \alpha_{k'}) - D\Psi(\sum_{k'} \alpha_{k'})} \quad (5)$$

$$\beta^{new} = \beta \frac{\sum_k \sum_v \Psi(N_{k,v}+\beta) - D\Psi(\sum_{k'} \alpha_{k'})}{V \sum_k \Psi(N_k + \beta V) - KV\Psi(\beta V)} \quad (6)$$

ここで$\Psi(\cdot)$はディガンマ関数を示す．

３．実験

３．１．実験準備

ここでは，2.4 節で述べた LDA を『上井覚兼日記』に対して適用し，人名に関するトピック，つまりは人物グループを検出する．さらに，時間・空間を軸に人物グループの変化を分析していく．[3]による手法により『上井覚兼日記』から人名を取得し，人名リポジトリに登録した．本実験ではその人名データを用いて実験を進めた．人名の異なり数は 520，述べ 4,025 だった．同様に，地名データは地名リポジトリから取得した．地名データは日記全体ではなく，天正 10 年 11 月から天正 12 年 4 月までの期間のみが登録されている．これらにより人物関係データを構成する．

LDA におけるパラメータとして，トピック数を 10，図 6 で示した崩壊型ギブスサンプリングの各繰り返しにおいてパープレキシティ（Perplexity）を計測しており，100 回目あたりで急速にその数値が減少し，900 回あたりで落ち着いた．そのため，その回数を 1000 とした．このサンプリングまたハイパーパラメータであるαおよびβの各要素の初期値は 0.1 とした．

３．２．トピックとその時系列変化

図 7 に LDA により検出されたトピックとそれぞれに属する人名（出現頻度順）を示す．ここでは史料上に現れる“拙者”は日記の書き手である上井覚兼として処理した．各トピックを端的に説明するのは困難であるため断定できないが，トピック 1 は島津家家老や肥後に関わる人物，トピック 2 は大隅関係者や島津家家老，トピック 3 は龍造寺氏・肥前・筑前の関係者およびその情勢に関わる人物，トピック 4 は上井覚兼の一門や宮崎衆，

©2017 Information Processing Society of Japan

(a) 宮崎　(b) 紫波州崎

(c) 佐土原　(d) 鹿児島

(e) 八代　(f) 有馬

図 10　トピックの時空間変化
Figure 10　Spatio-temporal change of topics

トピック 5 は島津歳久ら，トピック 8 は島津義久ら，トピック 9 は島津義弘ら，トピック 10 は島津家久らであろうと思われる．トピック 2 と 8 は非常によく似ているように思えるが，8 は重臣・家老に相当する人物が上がっているのに対し，2 では大隅・肝付関係者が並んでいるため，上記のように分類できると考えている．

図 8 は LDA によるトピックの検出結果を時系列に並べたものである．この結果よりいずれのトピックも時系的変化が認められた．大隅関係者や島津家家老に関係するトピックであろうと考えられるトピック 2 は天正 2 年ごろに頻出している．このとき上井覚兼は内城の奏者である．図 9 はトピック 2 に属する人名を時系列に並べた結

©2017 Information Processing Society of Japan

果を示す. 天正初年は島津氏による大隅攻略の進展により, 内城の奏者である上井覚兼が外交的にも内政的にも島津家老臣や大隅関係者と共に奮闘していた時期と言えるため, 実験結果と合致する．天正 10 年以降はその頻度が下がるものの，当主義久，弟である義弘，家老，他家の当主・重臣が属している．沖田畷の戦い（天正 10 年）や戸次川の戦い（天正 14 年）のあたりではその関係者が属していることがわかる.

トピック 9 は天正 12 年 10 月−12 月で最も頻出しており，天正 13 年 7 月−9 月まで高い頻度が続く. これは島津家が肥後の攻略を進めている時期である. トピック 9 は島津義弘を中心とした人物グループとしてみることができるため, この間は義弘を中心に奮闘していたことが伺える. トピック 10 は天正 13 年 7 月−9 月以降，天正 14 年 7 月−9 月まで頻出する．家久を中心とした人物グループと思われるトピック 10 は家久，さらに義弘や島津忠長などとともに, 筑後さらに筑前と北上していく時期であり，その動きと合致している.

３．３．空間とトピックの関係

図 10 に，天正 10 年 11 月から天正 12 年 4 月までの間での, 地名ごとのトピックの時系列変化を示す．ここでは，この間で本文に出現する地名のうち頻度順に上位 6 ヶ所（宮崎，紫波州崎，佐土原，鹿児島，八代，有馬）を選定した.

宮崎は上井覚兼が城主となった地域である. この間では頻度順にトピック 1，4，2，6 の順だった．トピック 1 と 4 および 6 は同じ時期に出現していた．他方，トピック 8 は，その頻度は低いものの，沖田畷の戦いの時期である天正 12 年 2-3 月に突然出現している.

紫波州崎は宮崎市折生にある城山を指すと考えられる. 宮崎から近い位置にあることから宮崎と同様の結果であろうと予想していたが, 様相は異なっていた．とくに天正 11 年 8 月付近および天正 12 年 1 月付近で顕著に異なっていることがわかった. これらの時期に上井覚兼は半月ほど鹿児島に滞在していることがわかっているが, その影響があるかもしれない.

この間の佐土原は島津家久の居城である. しかしながら, 天正 11 年 4 月−6 月以外ではあまり頻出しなかった. 家久自身が佐土原にいる時期があまりなかったためかもしれない. 鹿児島は当主義久の居城であることからさまざまな人物が登場している．最も頻出したのはトピック 1 だった.

八代および有馬は上井覚兼の居城とは九州の反対側に位置する．それらの位置から天正 12 年 3 月の沖田畷の戦い付近で頻出すると予想していたが, 実際には天正 10-11 年の方が圧倒的に頻出していることがわかった．八代において，他の地域に比べトピック 10 の割合が大きく, 天正 11 年 9 月にトピック 10 が頻出している．天正 11 年 8 月にトピック 1 が頻出していることもあり, なんらかの関係がありえるかもしれない. 有馬においてもトピック 10 の割合は大きいが，八代とは異なり, 天正 11 年 11-12 月に頻出している. また，他の地域に比べトピック 5 の割合が大きい. 特に天正 12 年 3 月に頻出している.

４．おわりに

本論文では天正期古記録『上井覚兼日記』を例に，史料本文に登場する人名を対象に LDA を用いた人物グループを検出の手法を述べた. さらに, 時間的・空間的変化によるトピックの変動について示した. 本研究の手法の検証から, LDA による分析手法の適用可能性およびその有効性を確認することができた.

LDA は対象とする史料の分類に役立つだけでなく，時系列変化や空間的変化により，検出したトピックの変動をつぶさに確認していくことが可能である．また，もとの史料に分析結果を照らし合わせながら分析していくことも可能である. また, LDA は教師なし学習・ベイズ学習に分類される機械学習の手法であることから, 適用範囲の柔軟性に優れる．さらに，分析結果を分類指標として扱えば, deep learning, ニューラルネットワークのような教師あり学習による分析にもつながっていくと考えている．そのため LDA だけでなくさらなる分析手法へと適用していき, 史料の理解につなげていくことを考えている.

『上井覚兼日記』においては，まだ人名・地名の全てのデータを取得しきれていない. そのためデータの網羅性に努めていく必要があると考えている．図 9 をみると“島津以久”“以久”のように人名統合がまだ行われていない. これは地名においても同様である. 今後データの洗練に努めていきたい.

これまで『上井覚兼日記』のみで分析を進めてきた．しかしながら『島津家文書』のように関係する史料が存在しており, これらを踏まえさらなる分析を進めていく予定である．これにより，上井覚兼や天正期の薩摩島津家に関する理解, さらには『上井覚兼日記』自体の深い理解につながっていくと考えている.

謝辞

本研究の成果の一部は, JSPS 科研費 JP26730167, JP15H01722, JP2624004, JP26220402, JP17H00773, JP16H01897，JP15H01723 の助成を受けたものによる.

©2017 Information Processing Society of Japan

参考文献

1) 東京大学史料編纂所：事業〈http://www.hi.u-tokyo.ac.jp/about_hi/mission-j.html〉（参照 2017-09-05）.
2) 山田太造 , 遠藤珠紀 , 荒木裕行 , 井上聡 , 久留島 典子：前近代日本史史料から人名を集める，じんもんこん 2016 論文集，pp.159-164（2016）.
3) 山田太造，野村朋弘，井上聡「トピックモデルを用いた天正期古記録『上井覚兼日記』における人物間関係の検出」，じんもんこん 2014 論文集，vol.2014，no.3，pp.131-138（2014）.
4) 山田太造，井上聡，遠藤珠紀，久留島典子：日本史史料読解支援のための候補文字検索，じんもんこん 2011 論文集，Vol.2011，No.8，pp.43-50（2011）.
5) D. M. Blei, A. Y. Ng, and M. I. Jordan: “Latent Dirichlet Allocation,” Journal of Machine Learning Research, vol.3, pp.993-1022（2003）.
6) T.L.Griffiths and M.Steyvers: Finding scientific topics, Proc. of the National Academy of Sciences of the United States of America, vol.101, pp.5228-5235(2004).
7) H. M Wallach, D. M Mimno, A. McCallum: Rethinking LDA: Why priors matter. proc of NIPS2009, pp.1973-1981(2009).
8) ネットアドバンス：JapanKnowledge Lib，入手先〈http://japanknowledge.com/library/〉（参照 2017-11-15）.
9) コトバンク，入手先〈https://kotobank.jp/〉（参照 2017-11-15）.
10) 国立歴史民俗博物館：地域蘭学者門人帳データベース，入手先〈http://www.rekihaku.ac.jp/doc/gaiyou/rung.html〉（参照 2017-11-15）.
11) 国立歴史民俗博物館：江戸商人・職人データベース，入手先〈http://www.rekihaku.ac.jp/doc/gaiyou/edos.html〉（参照 2017-11-15）.
12) 国文学研究資料館：歴史人物画像データベース，入手先〈http://base1.nijl.ac.jp/~rekijin/〉（参照 2017-11-15）.
13) 国立歴史民俗博物館：日本荘園，入手先〈https://www.rekihaku.ac.jp/doc/gaiyou/syouen.html〉（参照 2017-11-15）.
14) 国立歴史民俗博物館：旧高旧領取調帳，入手先〈https://www.rekihaku.ac.jp/doc/gaiyou/kyuudaka.html〉（参照 2017-11-15）.

©2017 Information Processing Society of Japan