Songrium 派生要因分析：N 次創作活動のモデル化による派生要因鑑賞サービス

佃　洸摂† 　石田　啓介† 　濱崎　雅弘† 　後藤　真孝†

† 産業技術総合研究所　〒 305–8568 茨城県つくば市梅園 1-1-1
E-mail: †{k.tsukuda,ksuke-ishida,masahiro.hamasaki,m.goto}@aist.go.jp

あらまし　我々はこれまで，オリジナルコンテンツから次々と新しい派生コンテンツが創作される N 次創作活動において，派生コンテンツの創作を引き起こした要因を推定するためのモデルを提案してきた．提案モデルでは，クリエータが派生コンテンツを創作する際に次の 3 つの要因が影響すると考える：（1）オリジナルコンテンツの魅力，（2）オリジナルコンテンツの人気，（3）派生コンテンツの人気．本稿では，動画共有サービスに投稿されたオリジナルコンテンツおよび派生コンテンツを対象として提案モデルを適用した，派生要因鑑賞サービス「Songrium 派生要因分析」について述べる．提案サービスでは，N 次創作活動におけるオリジナルコンテンツの特性の可視化，オリジナルコンテンツから派生コンテンツが創作される過程の可視化，クリエータが 3 つの各要因から受けた影響の大きさの可視化などの機能を提供する．動画コンテンツの探索や視聴をする際に，これまでは明らかではなかった，派生コンテンツ間の関係やクリエータの特性を可視化することで，ユーザやクリエータの新たなコンテンツとの出会いや N 次創作現象のより深い理解の支援を目指す．

キーワード　ユーザ生成コンテンツ，N 次創作，Web サービス

1. はじめに

プロのクリエータが中心となってコンテンツを創作し Web 上で発信していた時代は過ぎ，現在ではアマチュアのクリエータも容易にコンテンツの創作と Web 上での発信ができるようになった．そのようなコンテンツは UGC（User Generated Content）と呼ばれる．一から新しいコンテンツを創作することが難しいクリエータの間では，既存のオリジナルコンテンツを元にした新しい派生コンテンツの創作が盛んに行われている．たとえば YouTube(注1) やニコニコ動画 (注2) では，既存の楽曲に合わせて歌ったり踊ったりする派生動画が多数投稿されている [1] [2] [3]．他にも，3D プリンタ用の 3D モデルを共有できる web サービス Thigiverse(注3) では，あるクリエータが投稿したオリジナルのモデルを，別のクリエータがダウンロードして加工した新たなモデルの投稿を積極的に促すようにサービスが設計されている [4]．このような創作活動では，第 1 世代にあたるオリジナルコンテンツから第 2 世代のコンテンツが創作され，第 2 世代のコンテンツに影響されたクリエータが第 3 世代のコンテンツを，第 N 世代のコンテンツに影響されたクリエータが第 N+1 世代のコンテンツを創作することが起きている．このような派生コンテンツの創作活動は「N 次創作活動 [5]」と呼ばれる．

N 次創作活動は盛んであるが，それぞれの派生コンテンツはなぜ創作されたのだろうか．創作を引き起こした要因が Web 上で明示的に記述されることは少ないため，要因の特定は難しい．我々はこれまでに，クリエータが派生コンテンツを創作する際には，（1）オリジナルコンテンツの魅力，（2）オリジナルコンテンツの人気，（3）派生コンテンツの人気，の 3 つの要因が影響するという仮定に基づき，派生コンテンツの創作を引き起こした要因を推定するモデルを提案してきた [6]．3 つの各要因に対する重要度はクリエータによって異なると考えられるため，提案モデルではクリエータと各要因の関係の強さも考慮する．さらに提案モデルでは，オリジナルコンテンツと派生コンテンツの人気度を考慮するためにコンテンツのランキング情報を利用する．具体的には，Web の検索結果の閲覧モデル [7] を参考に，ランキングで上位に出現したコンテンツほど高い確率で多くのクリエータの目に触れるため N 次創作活動において高い影響力を持つという仮説に基づいたモデル化を行った．我々は音楽コンテンツの N 次創作活動データを用いたこれまでの実験 [6] で，上記の 3 つの要因を考慮した提案モデルの有用性を定量的に示した．また，コンテンツの人気度に基づくランキングを考慮する際は，ランキング閲覧時のクリエータの振る舞いを反映することでモデル化の精度が高くなることも示した．

本研究では，動画共有サービス「ニコニコ動画」に投稿された，VOCALOID [8] に関するオリジナルコンテンツおよび派生コンテンツを対象として提案モデルを適用した．その上で，誰もが Web ブラウザ上で結果を閲覧できるサービス「Songrium

(注1)：http://www.youtube.com
(注2)：http://www.nicovideo.jp
(注3)：http://www.thingiverse.com

派生要因分析 (注4)」を実現して提供する．提案サービスでは，N 次創作活動におけるオリジナルコンテンツの特性の可視化，オリジナルコンテンツから派生コンテンツが創作される過程の可視化，クリエータが 3 つの各要因から受けた影響の大きさの可視化などの機能を提供する．動画コンテンツの探索や視聴をする際に，これまでは明らかではなかった，派生コンテンツ間の関係やクリエータの特性を可視化することで，ユーザやクリエータの新たなコンテンツとの出会いや N 次創作現象のより深い理解の支援を目指す．

2. 関 連 研 究

[N 次創作活動の分析] Eto ら [9] は 3D モデルの創作と共有および，既存モデルを用いた派生モデルの創作を可能とする Web サイトを公開した．彼らは 10.4%のモデルが他のモデルの派生元になっていること，派生創作活動は最長で第 4 世代まで続いたことを報告している．Cheliotic ら [10] はオンライン音楽コミュニティの ccMixter(注5) における音楽コンテンツの N 次創作活動の調査を行った．調査の結果，N 次創作活動によってコミュニティ全体のコンテンツ量が飛躍的に増加すると共に，コンテンツの多様性も高くなることが明らかになった．Hamasaki ら [1] はニコニコ動画上での派生創作活動を対象とし，オリジナルコンテンツと派生コンテンツの間の自明な親子関係を扱い，各オリジナルコンテンツの派生コンテンツ数などの統計量を分析した．以上の研究では派生コンテンツの創作が「どのように」行われたを分析しているのに対して，我々は「なぜ」派生コンテンツが創作されたかに着目し，その要因の推定結果に基づく Web サービスを実現する．

[社会活動における影響のモデル化] 社会活動でのユーザ間の影響が推定できると，影響力の大きいユーザの特定 [11] や商品の推薦 [12] において有用であるため，多くの推定手法が提案されてきた．主要なアプローチのひとつは，独立カスケードモデル [13] などの情報拡散モデルを用いることである．このモデルでは離散時間を仮定しているが，Saito ら [14] は連続時間を扱うためにポアソン過程に基づくモデルを提案した．彼らのモデルでは，ユーザ間の影響の有無を表すグラフ情報が必要であるが，Iwata ら [15] はグラフ情報が与えられなくてもユーザ間の影響関係を推定するモデル，Shared Cascade Poisson Process（SCPP）を提案した．Tanaka ら [16] は SCPP を拡張し，商品の購買を引き起こした要因を推定するモデルを提案した．彼らはユーザ間の影響に加えて，ユーザがメディア広告から受ける影響も考慮し，SCPP の考えが商品の購買行動のモデル化にも有用であることを示した．我々のモデルも SCPP および Tanaka ら [16] のモデルを拡張したものであるが，次の点で既存モデルと異なる：（1）既存モデルではユーザによって選択されたアイテム（購入された商品など）の影響は考慮する必要がなかった．しかし，N 次創作活動においては，クリエータによって創作された派生コンテンツ自体も他のクリエータの創作活動に影響を与える．そのため我々はオリジナルコンテンツと派生コンテンツの両方の影響を扱えるように SCPP を拡張した．（2）既存モデルではアイテムの人気度がユーザのアイテム選択に与える影響は時間によらず一定であると仮定していたが，我々はその影響は時間によって変化すると仮定する．我々の提案モデルではコンテンツの人気度ランキングとクリエータのランキング閲覧行動を考慮することで，オリジナルコンテンツおよび派生コンテンツの時間に応じた人気度を扱う．

(注4)：http://factor.songrium.jp
(注5)：http://ccmixter.org

3. モ デ ル

社会活動をモデル化する際は，アイテムに対するユーザの好み（本研究ではオリジナルコンテンツの魅力に対応）およびユーザ間の影響を考慮するのが一般的である [15], [16]．しかし，N 次創作活動においては，クリエータ間の影響が派生コンテンツの創作を引き起こす要因になっていることは報告はされていない [1], [9], [10]．その代わりに，人気の高いコンテンツほど多くの派生コンテンツが作成される，「rich-get-richer [17]」に基づく現象が存在するであろうことがこれまでの報告から考えられる [1]．この考えに基づき，我々はこれまでに，オリジナルコンテンツの魅力，オリジナルコンテンツの人気度，および派生コンテンツの人気度が N 次創作活動において重要な役割を果たすと仮定し，派生コンテンツの創作を引き起こした要因を推定するモデルを提案した [6]．また，上記の 3 要因のみを考慮したモデルの方が，3 要因に加えてユーザ間の影響も考慮したモデルよりも優れたモデルであることを実世界のデータを用いて定量的に示した [6]．本章では，我々の提案モデルについて述べる．

3.1 記号の定義

「椅子の 3D モデル」や「既存楽曲を歌った動画」などのカテゴリおよびデータの観測期間 T が与えられたとき，$\mathcal{I}$ をある Web サービスに時刻 0 から T の間に投稿された対象カテゴリ内のオリジナルコンテンツの集合とする．(t^p_{ij}, u^p_{ij}) をオリジナルコンテンツ $i \in \mathcal{I}$ の j 番目に投稿された派生コンテンツとする．具体的には，クリエータが $u^p_{ij} \in \mathcal{U}$ が時刻 t^p_{ij} に i の派生コンテンツを投稿したことを表す．$\mathcal{U}$ は全クリエータ集合である．一般性を失うことなく，派生コンテンツの投稿イベントは投稿時刻に関して昇順に並べられているとする（$j < j'$ に対して $t^p_{ij} \leqq t^p_{ij'}$）．J_i を観測期間中に投稿された i の全派生コンテンツ数とすると，i の派生コンテンツの投稿イベント集合は $\mathcal{D}_i = \{(t^p_{ij}, u^p_{ij})\}_{j=1}^{J_i}$ で表される．従って，全てのオリジナルコンテンツの派生コンテンツの投稿イベント集合は $\mathcal{D} = \{\mathcal{D}_i\}_{i \in \mathcal{I}}$ と表される．

クリエータは Web サービス上でオリジナルコンテンツの人気度に基づくランキングを閲覧できるとする．(t^o_{ik}, r^o_{ik}) を $i \in \mathcal{I}$ の k 番目のランキング入りイベントとし，i が時刻 t^o_{ik} に r^o_{ik} 位にランキング入りしたことを表す．一般性を失うことなく，ランキング入りイベントはランキング入りした時刻に関して昇順に並べられているとする．K^o_i を i が観測期間においてランキング入りした回数とすると，i のランキング入りイベントの集合は $\mathcal{O}_i = \{(t^o_{ik}, r^o_{ik})\}_{k=1}^{K^o_i}$ と表され，全オリジナルコンテンツのランキング入りイベント集合は $\mathcal{O} = \{\mathcal{O}_i\}_{i \in \mathcal{I}}$ と表される．

図 1　クリエータ u が時刻 t にオリジナルコンテンツ i の派生コンテンツを投稿するレート．

クリエータは派生コンテンツの人気度に基づくランキングも閲覧できるとする．オリジナルコンテンツの場合と同様，(t^c_{ik}, r^c_{ik}) を i の派生コンテンツの k 番目のランキング入りイベントとする．K^c_i を i の派生コンテンツが観測期間においてランキング入りした回数とすると，i の派生コンテンツのランキング入りイベントの集合は $\mathcal{C}_i = \{(t^c_{ik}, r^c_{ik})\}_{k=1}^{K^c_i}$ と表される．$\mathcal{C}_i$ には i の全ての派生コンテンツのランキング入りイベントが含まれており，(t^c_{ik}, r^c_{ik}) と $(t^c_{ik'}, r^c_{ik'})$ は異なる派生コンテンツのランキング入りイベントとなる可能性もある．全オリジナルコンテンツの全派生コンテンツのランキング入りイベント集合は $\mathcal{C} = \{\mathcal{C}_i\}_{i \in \mathcal{I}}$ と表される．

3.2　要　　因

3.2.1　オリジナルコンテンツの魅力

オリジナルコンテンツ i の人気が低い場合でも，クリエータ u が i を魅力的だと感じれば，u は i の派生コンテンツを創作するかもしれない．i の魅力は i の様々な特徴量から引き起こされうる．たとえば楽曲の場合，メロディや歌詞などの特徴量が考えられる．オリジナルコンテンツの魅力から受ける影響の大きさはクリエータによって異なり，またその影響に基づいて派生コンテンツを投稿するレートは時刻 0 から T の間で一定であると仮定する（図 1（a））．ここで，時刻 t におけるレートとは，クリエータが時刻 t に i の派生コンテンツを投稿する確率を表す．この仮定に基づき，i の魅力が要因となってクリエータ u が i の派生コンテンツを投稿するレートを次式で求める．

$$f_i(u) = \alpha_i \theta_{0u}. \tag{1}$$

$\alpha_i \geqq 0$ はオリジナルコンテンツ i の魅力を表す．$\theta_{0u} \geqq 0$ は u がオリジナルコンテンツの魅力に影響を受ける度合いであり，$\sum_{u \in \mathcal{U}} \theta_{0u} = 1$ を満たす．図 1（a）では，青いバーの高さが $\alpha_i \theta_{0u}$ に相当する．

3.2.2　オリジナルコンテンツの人気

オリジナルコンテンツ i が消費者の間で人気であれば，クリエータ u が「このオリジナルコンテンツの派生コンテンツを創作すれば，自分の派生コンテンツも沢山の人に見てもらえるかもしれない」と考えて i の派生コンテンツを創作する，という状況が考えられる．3.1 節で述べたように，本研究ではクリエータがオリジナルコンテンツの人気度ランキングを閲覧できると仮定する．2 つのコンテンツがランキング入りしているとき，ランキング上位のコンテンツの方が高い影響力を持つという仮説を立てる．また，オリジナルコンテンツの人気が与える影響はクリエータによって異なり，影響力は時間の経過と共に減衰すると仮定する．この仮定に基づき，i の人気の影響で u が時刻 t に i の派生コンテンツを投稿するレートを次式で求める．

$$h_{o(i,t',r')}(t,u) = \begin{cases} rb(r')\omega_i \theta_{-1u} \mathrm{e}^{-\gamma_o(t-t')} & \text{if} \quad t' < t \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

r' は i の時刻 t' におけるランキング内での順位を，関数 rb は順位によるバイアスを表す．ユーザの Web 検索結果の閲覧時の振る舞いを分析した研究では，Web ページの順位が低くなると共に，閲覧される確率は急激に下がることが報告されている [7]．この閲覧モデルに基づき，提案モデルでは順位によるバイアスを $rb(r') = \frac{1}{r'}$ により求める．$\omega_i \geqq 0$ は i の人気による影響力を，$\theta_{-1u} \geqq 0$ は u がオリジナルコンテンツの人気から影響を受ける度合いを表し，$\sum_{u \in \mathcal{U}} \theta_{-1u} = 1$ を満たす．$\mathrm{e}^{-\gamma_o(t-t')}$ は，パラメータ $\gamma_o \geqq 0$ のもとで時間の経過と共に影響力が減衰することを表す．

図 1（b）では，i はランキングに 4 回出現している．最初のランキング入り時の順位を r'，時刻を t^o_{i1} とすると，t^o_{i1} における影響力は $rb(r')\omega_i \theta_{1u}$（図中の h_1 に相当）であり，時間の経過と共に影響力は減少していく．

3.2.3　派生コンテンツの人気

クリエータ u' が創作した i の派生コンテンツが消費者の間で人気であれば，その派生コンテンツを見たクリエータ u は，「この派生コンテンツの元になっているオリジナルの派生コンテンツを創作すれば，自分の派生コンテンツも沢山の人に見てもらえるかもしれない」と考えて i の派生コンテンツを創作する，という状況が考えられる．3.2.2 項で述べた仮説および仮定と同様に，i の派生コンテンツが時刻 t' に r' 位にランキング入りした場合，その影響で u が時刻 t に i の派生コンテンツを投稿するレートを次式で求める．

$$h_{d(i,t',r')}(t,u) = \begin{cases} rb(r')\sigma_i \theta_{-2u} \mathrm{e}^{-\gamma_d(t-t')} & \text{if} \quad t' < t \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

$\sigma_i \geqq 0$ は i の派生コンテンツの人気による影響力を，$\theta_{-2u} \geqq 0$ は u が派生コンテンツの人気から影響を受ける度合いを表し，$\sum_{u \in \mathcal{U}} \theta_{-2u} = 1$ を満たす．$\mathrm{e}^{-\gamma_d(t-t')}$ は，$\gamma_d \geqq 0$ のもとで時間の経過による影響力の減衰を表す．

図 1（c）および（d）では，i の 1 番目の派生コンテンツと

２番目の派生コンテンツの影響力をそれぞれ表している．図 1（c）の最初のランキング入りの順位を r' とすると，最初にランキング入りした時刻 t^c_{i1} における影響力は $rb(r')\sigma_i\theta_{2u}$（図中の h_2 に相当）であり，時間の経過と共に影響力は減少していく．

3.3 派生コンテンツの投稿レート

3.2.1 項から 3.2.3 項で述べた要因に基づき，クリエータ u が時刻 t にオリジナルコンテンツ i の派生コンテンツを投稿するレートを次式で表す．

$$\lambda_i(t,u) = f_i(u) + \sum_{(t',r')\in\mathcal{O}_{it}} h_{o(i,t',r')}(t,u) + \sum_{(t',r')\in\mathcal{C}_{it}} h_{d(i,t',r')}(t,u). \quad (4)$$

ここで，$\mathcal{O}_{it} = \{(t',r')|(t',r') \in \mathcal{O}_i \text{ and } t' < t\}$，$\mathcal{C}_{it} = \{(t',r')|(t',r') \in \mathcal{C}_i \text{ and } t' < t\}$ である．$\lambda_i(t,u)$ は図 1（e）中の h_3 に相当する．

4. パラメータ推定

観測データ $\mathcal{D}$，$\mathcal{O}$，$\mathcal{C}$ が与えられたとき，$\mathcal{D}$ の尤度関数は次式で表される．

$$P(\mathcal{D}|\mathcal{O},\mathcal{C},\boldsymbol{\alpha},\boldsymbol{\omega},\boldsymbol{\sigma},\boldsymbol{\Theta},\boldsymbol{\gamma}) = \prod_{i\in\mathcal{I}} \exp\left(-\int_0^T \sum_{u\in\mathcal{U}} \lambda_i(t,u)dt\right) \prod_{j=1}^{J_i} \lambda_i(t^p_{ij}, t^p_{ij}). \quad (5)$$

パラメータは $\boldsymbol{\alpha} = \{\alpha_i\}_{i\in\mathcal{I}}$，$\boldsymbol{\omega} = \{\omega_i\}_{i\in\mathcal{I}}$，$\boldsymbol{\sigma} = \{\sigma_i\}_{i\in\mathcal{I}}$，$\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{u'}\}_{u'\in\mathcal{U}_+}$，$\boldsymbol{\theta}_{u'} = \{\theta_{u'u}\}_{u\in\mathcal{U}}$，$\boldsymbol{\gamma} = \{\gamma_o, \gamma_d\}$ である．$\mathcal{U}_+ = \{0,-1,-2\}$ であり，0，−1，−2 はそれぞれ，オリジナルコンテンツの魅力，オリジナルコンテンツの人気，派生コンテンツの人気，に相当する仮想クリエータを表す．式（5）の指数で表される項は，観測期間中に派生コンテンツが投稿されない確率を表す．また，式中の積分は次式のように計算される．

$$\int_0^T \sum_{u\in\mathcal{U}} \lambda_i(t,u)dt = \alpha_i T + \frac{\omega_i}{\gamma_o}\sum_{k=1}^{K_i^o} rb(r^o_{ik})\left(1-\mathrm{e}^{-\gamma_o(T-t^o_{ik})}\right) + \frac{\sigma_i}{\gamma_d}\sum_{k=1}^{K_i^c} rb(r^c_{ik})\left(1-\mathrm{e}^{-\gamma_d(T-t^c_{ik})}\right). \quad (6)$$

Iwata ら [15] に倣い，オリジナルコンテンツ i の j 番目の派生コンテンツの投稿を引き起こしたイベントのインデックスを表す潜在変数 $z_{ij} \in \{0,1,\cdots,|\mathcal{O}_{it}|+|\mathcal{C}_{it}|\}$ を導入する．$z_{ij} = 0$ は派生コンテンツの投稿がオリジナルコンテンツの魅力によって引き起こされたことを，$1 \leqq z_{ij} \leqq |\mathcal{O}_{it}|$ はオリジナルコンテンツの人気によって引き起こされたことを，$|\mathcal{O}_{it}|+1 \leqq z_{ij} \leqq |\mathcal{O}_{it}|+|\mathcal{C}_{it}|$ は派生コンテンツの人気によって引き起こされたことをそれぞれ表す．潜在変数を用いることで，式（4）は $\lambda_i(t,u) = \sum_z \lambda_i(t,u,z)$ と書ける．ここで，

$$\lambda_i(t,u,z) = \begin{cases} f_i(u) & \text{if} \quad z=0 \\ h_{o(i,t^o_{iz},r^o_{iz})}(t,u) & \text{if} \quad 1 \leqq z \leqq |\mathcal{O}_{it}| \\ h_{d(i,t^c_{iz'},r^c_{iz'})}(t,u) & \text{if} \quad |\mathcal{O}_{it\backslash u}|+1 \leqq z \leqq |\mathcal{O}_{it\backslash u}|+|\mathcal{C}_{it\backslash u}|. \end{cases} \quad (7)$$

表 1　データセットの統計量

	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{D}\rvert$	$\lvert\mathcal{C}\rvert$	$\lvert\mathcal{U}\rvert$
歌ってみた	5,632	157,086	402,406	152,086	31,782
踊ってみた	784	93,828	31,677	76,457	3,095
演奏してみた	658	95,024	9,316	52,514	1,066

であり，$z' = z - |\mathcal{O}_{it}|$ である．

式 (5)，(6)，(7) を用いることで，$\mathcal{D}$ および $\mathcal{Z} = \{\{z_{ij}\}_{j=1}^{J_i}\}_{i\in\mathcal{I}}$ の同時分布は次式で表される．

$$P(\mathcal{D},\mathcal{Z}|\mathcal{O},\mathcal{C},\boldsymbol{\alpha},\boldsymbol{\omega},\boldsymbol{\sigma},\boldsymbol{\Theta},\boldsymbol{\gamma}) = \prod_{i\in\mathcal{I}} \exp\left[\alpha_i T + \frac{\omega_i}{\gamma_o}\sum_{k=1}^{K_i^o} rb(r^o_{ik})\left(1-\mathrm{e}^{-\gamma_o(T-t^o_{ik})}\right) + \frac{\sigma_i}{\gamma_d}\sum_{k=1}^{K_i^c} rb(r^c_{ik})\left(1-\mathrm{e}^{-\gamma_d(T-t^c_{ik})}\right)\right] \prod_{j=1}^{J_i} \lambda_i(t^{\mathrm{p}}_{ij}, t^{\mathrm{p}}_{ij}, z_{ij}). \quad (8)$$

$\alpha_i \in \boldsymbol{\alpha}$ の事前分布としてガンマ分布を仮定すると，ハイパーパラメータ a および b を用いて $P(\alpha_i|a,b) \propto \alpha_i^{a-1}\exp(-b\alpha_i)$ と表される．Iwata ら [15] に倣い，提案モデルでは $a = b = 1$ とした．同様に $\omega_i \in \boldsymbol{\omega}$ および $\sigma_i \in \boldsymbol{\sigma}$ に対しても事前分布としてガンマ分布を仮定する．さらに，$\boldsymbol{\theta}_{u'} \in \boldsymbol{\Theta}$ の事前分布としてディリクレ分布を仮定すると，ハイパーパラメータ β を用いて $P(\boldsymbol{\theta}_{u'}|\beta) \propto \prod_{u\in\mathcal{U}} \theta_{u'u}^{\beta-1}$ と表される．以上のように事前分布を仮定することで，次式を基にパラメータ $\boldsymbol{\alpha}$，$\boldsymbol{\omega}$，$\boldsymbol{\sigma}$，$\boldsymbol{\Theta}$，および $\boldsymbol{\gamma}$ の積分消去が可能になる．

$$P(\mathcal{D},\mathcal{Z}|\mathcal{O},\mathcal{C},\boldsymbol{\gamma},\beta,a,b) = \iiiint P(\mathcal{D},\mathcal{Z}|\mathcal{O},\mathcal{C},\boldsymbol{\alpha},\boldsymbol{\omega},\boldsymbol{\sigma},\boldsymbol{\Theta},\boldsymbol{\gamma})P(\boldsymbol{\alpha}|a,b) \times P(\boldsymbol{\omega}|a,b)P(\boldsymbol{\sigma}|a,b)P(\boldsymbol{\Theta}|\beta)d\boldsymbol{\alpha}d\boldsymbol{\omega}d\boldsymbol{\sigma}d\boldsymbol{\Theta}. \quad (9)$$

式 (9) に基づき，確率的 EM アルゴリズムを用いて $z_{ij} \in \mathcal{Z}$ のサンプリングを行う．具体的には，E ステップにおいて，z_{ij} 以外の全ての潜在変数の割り当てが与えられた下で，z_{ij} の新たな値を以下の式に従いサンプリングする．

$$P(z_{ij} = y|\mathcal{D},\mathcal{Z}_{\backslash ij},\mathcal{O},\mathcal{C},\boldsymbol{\gamma},\beta,a,b) \propto \frac{P\left(\mathcal{D},\mathcal{Z}_{\backslash ij}, z_{ij}=y|\mathcal{O},\mathcal{C},\boldsymbol{\gamma},\beta,a,b\right)}{P\left(\mathcal{D}_{\backslash ij},\mathcal{Z}_{\backslash ij}|\mathcal{O},\mathcal{C},\boldsymbol{\gamma},\beta,a,b\right)}. \quad (10)$$

ここで，$y \in \{0,1,\cdots,|\mathcal{O}_{it}|+|\mathcal{C}_{it}|\}$ であり，$\backslash ij$ はオリジナルコンテンツ i の j 番目の派生コンテンツを除くことを意味する．M ステップでは，ニュートン法を用いて $\boldsymbol{\gamma}$ の値を更新する．最後に，十分な回数のイテレーションの後サンプリングされた $\mathcal{Z}$ を用いて，積分消去したパラメータの値を推定する．

5. Songrium 派生要因分析

本章では，派生要因鑑賞サービス「Songrium 派生要因分析」の機能について述べる．

5.1 データセット

Songrium 派生要因分析では，動画共有サービス「ニコニコ動画」の N 次創作活動データを使用した．ニコニコ動画では，

図 2　オリジナル動画のマッピング機能.

図 3　軸および各象限の説明.

特に音楽動画の N 次創作活動が盛んであり，2017 年 2 月の時点で 15 万件以上のオリジナル動画と 62 万件以上の派生動画が投稿されている．大部分のオリジナル動画は VOCALOID [8] と呼ばれる歌声合成技術を用いて創作された楽曲である．派生動画に関しては，オリジナル曲を歌う「歌ってみた」，オリジナル曲に合わせて踊る「踊ってみた」，オリジナル曲を演奏する「演奏してみた」の 3 つのカテゴリを対象とした．この 3 カテゴリは様々な派生動画カテゴリの中でも，ニコニコ動画において特に派生コンテンツの投稿が盛んに行われているカテゴリである．Songrium 派生要因分析では 2009 年 11 月 1 日から 2016 年 12 月 31 日の間に投稿されたオリジナル動画および派生動画を対象とした．いずれのカテゴリでも，対象期間において派生動画が 3 件未満のオリジナル動画および，派生動画の投稿件数が 4 件未満のクリエータはデータセットから除いた．

ニコニコ動画では，オリジナル動画および上記の 3 カテゴリの派生動画それぞれについて，日単位で人気動画の上位 100 件を見ることができる．1 日のランキングは，前日の閲覧数やコメント数などのいくつかの指標を元に作成される．我々は 2009 年 11 月 1 日から 2016 年 12 月 31 日の間の，オリジナル動画および 3 カテゴリの派生動画の上位 100 件のランキングデータを収集した．ランキングデータが日単位であるため，提案モデルでは派生動画の投稿時刻も日単位で扱った．

表 1 に Songrium 派生要因分析で使用したデータセットの統計量を示す．

5.2　オリジナルコンテンツのマッピング機能

Songrium 派生要因分析では，カテゴリごとにオリジナル動画が直交座標系の 2 次元空間上に表示される（図 2）．ひとつの円がひとつのオリジナル動画を表しており，派生動画の数に比例して円は大きく表示される．ユーザは画面左上に表示された 3 つのカテゴリのいずれかを選択することで，カテゴリを切り替えることができる．図 3 に示すように，縦軸はオリジナル動画の派生動画がコンテンツの魅力の影響を受けて創作されているか，コンテンツの人気の影響を受けて創作されているかを表す．横軸はオリジナル動画の派生動画を投稿したクリエータがコンテンツの魅力を重視しているか，コンテンツの人気を重視しているかを表す．このような 2 次元空間上にオリジナル動画をマッピングすることで，第一象限上のオリジナル動画は，コンテンツの人気を重視する多くのクリエータによって既に派生動画が創作されたオリジナル動画であり，派生創作における「定番曲」であると言える．第二象限には，動画のランキングには出現するが，人気があるからといって誰もが容易に派生動画を作れるとは限らないオリジナル動画が表示されるため，派生創作における「チャレンジ曲」であると言える．第三象限のオリジナル動画は，コンテンツの魅力を重視するクリエータが派生創作に使用した，ランキングにはあまり出現しないオリジナル動画であり，派生創作における「玄人受け曲」であると言える．第四象限には，動画のランキングにはまだそれほど出現していないが，コンテンツの人気を重視するクリエータが派生創作に使用したオリジナル動画が表示されるため，今後多くのクリエータによって派生動画が作成される可能性のある「流行候補曲」であると言える．このようにオリジナル曲をマッピングすることで，クリエータは自分がどのような曲の派生動画を作りたいかに応じて新しい観点からオリジナル動画を探すことが可能になる．以下では，オリジナル動画 i が与えられたときに，その座標を求める手法について述べる．

Algorithm 1 オリジナル動画 i の j 番目の派生動画の投稿に対する，コンテンツの魅力の影響度 E_{con} およびコンテンツの人気の影響度 E_{pop} の計算.

Require: $P(z_{ij}|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$

$E_{con} \Leftarrow P(z_{ij} = 0|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$, $E_{pop} \Leftarrow 0$
$y \Leftarrow 1$
while $y \leqq |\mathcal{O}_{it}| + |\mathcal{C}_{it}|$ **do**
　$E_{pop} \Leftarrow E_{pop} + P(z_{ij} = y|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$
　$y \Leftarrow y + 1$
end while
return E_{con}, E_{pop}

5.2.1 オリジナル動画の縦軸の値

式 (10) により計算される潜在変数の事後分布を用いることで，コンテンツの魅力およびコンテンツの人気が派生動画の投稿に影響を与えた度合いを求めることができる．Algorithm 1 に示す疑似コードでは，次の 2 つの値を求めている．ひとつは，オリジナル動画 i の j 番目の派生動画がコンテンツの魅力の影響で投稿された度合い，つまり 3.2.1 項の要因の影響で投稿された度合い E_{con}．もうひとつは，コンテンツの人気の影響で投稿された度合い，つまり 3.2.2 項または 3.2.3 項の要因の影響で投稿された度合い E_{pop} である．オリジナル動画 i の全ての派生動画について E_{con} と E_{pop} それぞれの和を計算することで，i の全派生動画におけるコンテンツの魅力の影響の大きさおよび，コンテンツの人気の影響の大きさを求めることができる．それぞれの和を $SUME_{con}$，$SUME_{pop}$ とし，縦軸の値が -1 から 1 の間に収まり，かつ値が大きいほど人気の影響が大きいことを表すために，次式によりオリジナル動画 i の縦軸の値 $V(i)$ を求める．

$$V(i) = 2\left(\frac{SUME_{pop}}{SUME_{pop} + SUME_{con}} - 0.5\right). \qquad (11)$$

5.2.2 オリジナル動画の横軸の値

クリエータ u が投稿した全ての派生動画が与えられたときに，各派生動画について Algorithm 1 により求められる E_{con} および E_{pop} それぞれの和 CE_{con} および CE_{pop} を計算することで，u がこれまでに派生動画を投稿する際にコンテンツの魅力および，コンテンツの人気を重視してきた度合いを求めることができる．オリジナル動画 i の全ての派生動画が与えられたときに，各派生動画を投稿したクリエータについて CE_{con} および CE_{pop} を求め，それぞれの和を計算する．これにより，i の派生動画を投稿したクリエータが普段派生コンテンツを投稿する際に，コンテンツの魅力とコンテンツの人気をどの程度重視しているかを求めることができる．オリジナル動画 i の派生動画を投稿したクリエータの CE_{con} および CE_{pop} の和をそれぞれ $CSUME_{con}$ および $CSUME_{pop}$ とし，横軸の値が -1 から 1 の間に収まり，かつ値が大きいほど人気を重視していることを表すために，次式によりオリジナル動画 i の横軸の値 $H(i)$ を求める．

$$H(i) = 2\left(\frac{CSUME_{pop}}{CSUME_{pop} + CSUME_{con}} - 0.5\right). \qquad (12)$$

5.3 派生動画の詳細表示機能

5.1 節で述べた 2 次元空間上のオリジナル動画をクリックして選択すると，その派生動画集合の詳細が表示される．図 4 に「歌ってみた」カテゴリのオリジナル動画「トリノコシティ(注6)」を選択した際に表示される結果を示す．表示される情報は大きく次の 3 種類から成る：(1) 派生動画ツリー，(2) 人気影響度推移グラフ，(3) クリエータ特性マップ．以下ではそれぞれの詳細を述べる．

(注6)：http://www.nicovideo.jp/watch/sm11559163

図 4 オリジナル動画を選択後に表示される，派生動画の詳細情報．

図 5 派生動画ツリー．

5.3.1 派生動画ツリー

派生動画ツリーでは，各派生動画がオリジナル動画またはいずれかの派生動画の影響を受けて創作された様子を可視化している．図 4 の派生動画ツリーを拡大したものを図 5 に示す．図中のひとつの円がひとつの派生動画を表している．横軸は動画の投稿日時であり，左端がオリジナル動画の投稿日時，右端が最も最近の派生動画の投稿日時に対応している．派生動画ツリーでは，オリジナル動画および派生動画がいくつかの層に分かれて表示される．最も上の層にはオリジナル動画のみが存在する．オリジナル動画の影響によって創作された派生動画は 1 次創作動画として上から 2 番目の層に表示される．さらに，1 次創作動画の影響によって創作された派生動画は 2 次創作動画として上から 3 番目の層に表示される．このとき，2 次創作の派生動画は，最も強い影響を受けた 1 次創作の派生動画と線で結ばれる．4 番目の層以降も同様である．円の大きさは，その派生動画に直接影響されて創作された派生動画の多さを表している．これまでにも，オリジナル動画の派生動画集合を可視化するサービスは存在したが [1]，提案モデルを使用することで，派生動画間の関係も把握することが可能になる．また，円をクリックして選択すると，対応する派生動画を視聴することができる．これにより，ユーザは N 次創作活動において大きな影響を与えた動画という新しい観点から動画を探して視聴することができる．さらに，図 6 に示すように，特定の円にマウスをホバーすることで，その派生動画の創作に直接的または間接的に影響を与えた派生動画および影響を与えられた派生動画のみが強調して表示される．こうすることで，ユーザは気になった派生動画に特化した派生動画ツリーを見ることができる．

オリジナル動画 i の派生動画ツリーを生成するためには，j 番

図 6　特定の派生動画にマウスをホバーした状態.

目の派生動画の創作に影響を与えた動画を特定できれば良い．そのために，式 (10) によって求められる z_{ij} の値の確率分布からサンプリングされた値 y を用いる．$y = 0$ または $1 \leqq y \leqq |O_{it}|$ であるとき，オリジナル動画の魅力またはオリジナル動画のランキング入りイベントの影響によって j 番目の派生動画が創作されたことを表しており，派生動画ツリー上ではオリジナル動画と線で結ばれた 1 次創作動画として表示される．y が j'（$j' < j$）番目の派生動画のランキング入りイベントに対応している場合，j' 番目の派生動画の影響によって j 番目の派生動画が創作されたことを表しており，派生動画ツリー上では j' 番目の派生動画と線で結ばれ，j' 番目の派生動画のひとつ下の階層に表示される．

5.3.2　人気影響度推移グラフ

人気影響度推移グラフでは，派生動画の創作に対してコンテンツの人気が与えた影響の大きさの 1ヶ月ごとの推移を可視化している．図 4 の人気影響度推移グラフを拡大したものを図 7 に示す．派生動画ツリー同様，横軸は動画の投稿日時であり，左端がオリジナル動画の投稿日時，右端が最も最近の派生動画の投稿日時に対応している．縦軸は，各月に投稿された派生動画が，コンテンツの人気に影響されて創作された度合いを表している．この値が小さい区間は，コンテンツの魅力に影響されて派生動画が創作された度合いが高いことを意味する．このグラフを見ることで，ユーザは選択したオリジナル動画の N 次創作活動における人気の影響力の移り変わりを容易に知ることができる．また，クリエータにとっては，そのオリジナル動画の派生動画を創作するか否かの意思決定の手段にもなりうる．たとえば，日頃からコンテンツの人気を重視して派生動画を創作しているクリエータであれば，グラフの右端における人気の影響力が大きいオリジナル動画を対象に派生動画を創作すれば良い．逆に，あえてコンテンツの人気の影響力が高くないオリジナル動画の派生動画を創作することを好むクリエータであれば，グラフの右端における人気の影響力が小さいオリジナル動画を見つけて派生動画を創作すれば良い．

オリジナル動画 i の人気影響度推移グラフを生成するためには，各月に投稿された派生動画集合における，コンテンツの人気の影響の度合いを求めれば良い．すなわち，ある月に投稿された派生動画集合に対して，5.2.1 項と同様にして $SUME_{con}$ および $SUME_{pop}$ を求め，$\frac{SUME_{pop}}{SUME_{con}+SUME_{pop}}$ を縦軸の値とする．

図 7　人気影響度推移グラフ.

図 8　クリエータ特性マップ.

5.3.3　クリエータ特性マップ

クリエータ特性マップでは，ユーザが選択したオリジナル動画の派生動画を投稿したクリエータが，3.2 節の 3 つの各要因を日頃どの程度重視しているかを可視化している．図 4 のクリエータ特性マップを拡大したものを図 8 に示す．ひとつの点が一人のクリエータに対応している．三角グラフの各頂点の「オリジナル魅力」「オリジナル人気」「派生人気」はそれぞれオリジナルコンテンツの魅力，オリジナルコンテンツの人気，派生コンテンツの人気を表す．たとえば図 8 では，オリジナル動画の魅力とオリジナル動画の人気を重視する多くのクリエータがこのオリジナル動画の派生動画を投稿していることがわかる．このような情報を提示することで，あるクリエータがこのオリジナル動画の派生動画を創作するか決める際に，自分が重視する要因に多くのクリエータが分布しているかを参考にするような，意思決定の支援につながる可能性がある．

クリエータ特性マップを生成するためには，各クリエータがこれまでに投稿してきた各派生動画における 3 要因の影響力を求められれば良い．クリエータ u がオリジナル動画 i の j 番目の派生動画を創作した際の 3 要因の影響力は Algorithm 2 に示す疑似コードにより求められる．擬似コード中の E_f，E_{h_o}，E_{h_d} はそれぞれオリジナル動画の魅力の影響力，オリジナル動画の人気の影響力，派生動画の人気の影響力を表す．クリエータ u が創作した全ての派生動画について E_f，E_{h_o}，E_{h_d} それぞれの和 $SUME_f$，$SUME_{h_o}$，$SUME_{h_d}$ を求め，これらの和が 1 になるように正規化することで u の三角グラフ上の座標を求めることができる．

5.4　クリエータの詳細表示機能

5.3.3 項で述べたクリエータ特性マップ内のクリエータをクリックして選択すると，そのクリエータに関する詳細が表示される．図 9 にクリエータ「山猫さなえ」を選択した際に表示さ

Algorithm 2 オリジナル動画 i の j 番目の派生動画の投稿に対する，3 つの各要因の影響度の計算．

Require: $P(z_{ij}|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$

$E_f \Leftarrow P(z_{ij} = 0|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$, $E_{h_o} \Leftarrow 0$, $E_{h_d} \Leftarrow 0$
$y \Leftarrow 1$
while $y \leqq |\mathcal{O}_{it}| + |\mathcal{C}_{it}|$ **do**
  if $y \leqq |\mathcal{O}_{it}|$ **then**
    $E_{h_o} \Leftarrow E_{h_o} + P(z_{ij} = y|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$
  else
    $E_{h_d} \Leftarrow E_{h_d} + P(z_{ij} = y|\mathcal{D}, \mathcal{Z}_{\setminus ij}, \mathcal{O}, \mathcal{C}, \boldsymbol{\gamma}, \beta, a, b)$
  end if
  $y \Leftarrow y + 1$
end while
return E_f, E_{h_o}, E_{h_d}

図 9　クリエータの詳細情報．

れる結果を示す．この機能では，クリエータがこれまでに投稿した派生動画の一覧を見ることができる．さらに，各派生動画に関して，「派生要因となった可能性の高い動画」と「派生影響を与えた動画」が表示されるため，どのような派生動画がオリジナル曲から影響を受けているかや，どのような派生動画に影響を与えたかといった情報をユーザは俯瞰することができる．各派生動画の「派生グラフを表示」をクリックすることで，その派生動画についての 5.3 節で述べた詳細情報を見ることができる．これによりユーザは，派生動画の創作におけるクリエータの位置づけという観点から，動画を探索することが可能になる．

6. まとめと今後の課題

本稿では，派生コンテンツの創作を引き起こした要因を推定するモデルを動画共有サービス「ニコニコ動画」の N 次創作データに適用し，Web ブラウザ上で結果を閲覧できるサービス「Songrium 派生要因分析」について述べた．提案サービスの様々な機能により，ユーザが新しい観点から動画を探索して視聴したり，クリエータが派生創作に使用するオリジナル動画を決める際の意思決定を支援したりできる可能性について議論した．今後は，Web サービス上でのユーザの操作ログを分析することで，派生創作活動において大きな影響を与えた動画は視聴されやすいかなど，ユーザの動画視聴に与える影響について分析する予定である．また，本サービスにより提供する情報が，クリエータが派生動画の創作に使用するオリジナル動画を決める際にどのように役立つかを分析していく予定である．

謝　　辞

本研究の一部は，文部科学省科学研究費補助金研究活動スタート支援（課題番号 15H06887），JST CREST 及び ACCEL の支援を受けた．

文　　献

[1] M. Hamasaki, H. Takeda and T. Nishimura: “Network analysis of massively collaborative creation of multimedia contents: Case study of hatsune miku videos on nico nico douga”, UXTV, pp. 165–168 (2008).
[2] C. Cayari: “The youtube effect: How youtube has provided new ways to consume, create, and share music”, International Journal of Education & the Arts, **12**, 6, pp. 1–28 (2011).
[3] L. A. Liikkanen and A. Salovaara: “Music on youtube: user engagement with traditional, user-appropriated and derivative videos”, Computers in Human Behavior, **50**, pp. 108–124 (2015).
[4] S. Papadimitriou and E. E. Papalexakis: “Towards laws of the 3d-printable design web”, WebSci, pp. 255–256 (2014).
[5] M. Goto: “Grand challenges in music information research”, Dagstuhl Follow-Ups: Multimodal Music Processing, **3**, pp. 217–225 (2012).
[6] K. Tsukuda, M. Hamasaki and M. Goto: “Why did you cover that song?: Modeling n-th order derivative creation with content popularity”, CIKM, pp. 2239–2244 (2016).
[7] T. Joachims, L. Granka, B. Pan, H. Hembrooke and G. Gay: “Accurately interpreting clickthrough data as implicit feedback”, SIGIR, pp. 154–161 (2005).
[8] H. Kenmochi and H. Ohshita: “Vocaloid - commercial singing synthesizer based on sample concatenation”, INTERSPEECH, pp. 4009–4010 (2007).
[9] K. Eto, M. Hamasaki, K. Watanabe, Y. Kawasaki and T. Nishimura: “Modulobe: A creation and sharing platform for articulated models with complex motion”, ACE, pp. 305–308 (2008).
[10] G. Cheliotis and J. Yew: “An analysis of the social structure of remix culture”, C&T, pp. 165–174 (2009).
[11] X. Song, Y. Chi, K. Hino and B. L. Tseng: “Information flow modeling based on diffusion rate for prediction and ranking”, WWW, pp. 191–200 (2007).
[12] X. Song, B. L. Tseng, C.-Y. Lin and M.-T. Sun: “Personalized recommendation driven by information flow”, SIGIR, pp. 509–516 (2006).
[13] J. Yang and J. Leskovec: “Modeling information diffusion in implicit networks”, ICDM, pp. 599–608 (2010).
[14] K. Saito, M. Kimura, K. Ohara and H. Motoda: “Learning continuous-time information diffusion model for social behavioral data analysis”, ACML, pp. 322–337 (2009).
[15] T. Iwata, A. Shah and Z. Ghahramani: “Discovering latent influence in online social activities via shared cascade poisson processes”, KDD, pp. 266–274 (2013).
[16] Y. Tanaka, T. Kurashima, Y. Fujiwara, T. Iwata and H. Sawada: “Inferring latent triggers of purchases with consideration of social effects and media advertisements”, WSDM, pp. 543–552 (2016).
[17] A.-L. Barabási and R. Albert: “Emergence of scaling in random networks”, Science, **286**, 5439, pp. 509–512 (1999).