

# 動画コンテンツデータセットの動画メタデータおよび タグ情報の解析 – ニコニコデータセットを用いて –

黒瀬 浩<sup>1,a)</sup> 山田 茂樹<sup>2</sup>

**概要:** 株式会社ドワンゴと国立情報学研究所によりニコニコ動画のメタデータがニコニコデータセットとして提供されている。そのデータセットの動画メタデータとタグ情報を解析し、タグをキーとした経路探索の影響について確認する。

**キーワード:** 動画メタ情報, フォークソノミータグ, タグルーティング, ニコニコデータセット

## An Analysis of Video Meta-data and Thier Folksonomy Tags – Using Niconico Dataset –

KUROSE HIROSHI<sup>1,a)</sup> YAMADA SHIGEKI<sup>2</sup>

**Abstract:** Viode meta-information of Nicovideo has provided by DWANGO Co., Ltd. and National Institute of Informatics (NII). This paper shows the analysis results of the statistical data of the meta-information of the videos and attached folksonomy tags on their video contents.

**Keywords:** Video meta-data, Folksonomies, Tag based routing, Niconico dataset

### 1. はじめに

国立情報学研究所は、各種データセットを情報学関連分野の研究者に提供している [1]。その中に動画投稿サイトの動画およびコメントの情報がある。株式会社ドワンゴと国立情報学研究所によりニコニコ動画<sup>\*1</sup>の 800 万を超える動画に関するデータがニコニコデータセットとして提供されている。ニコニコ動画は、動画画面上に視聴者がコメントを投稿でき、コンテンツ投稿者、視聴者間のコミュニケーションを行うユニークなインターフェースを提供している。

実サイトのデータを用いた解析結果は、情報流通基盤の

設計および運用に有益であるため提示する。近年、コンテンツを中心とした情報流通基盤の検討が行われているためタグの解析も行い、タグをキーとしたコンテンツ探索を指向する情報流通基盤への影響を確認する。

本稿では、2 節でデータセットの概要を、3 節でデータセットの解析結果を、4 節でタグをキーとしたコンテンツ流通基盤への影響を確認する。

### 2. データセット

ニコニコデータセットは、2007 年 3 月初旬から 2012 年 11 月初旬までにニコニコ動画に投稿された約 830 万件の動画メタデータ約 12GB と、それらの動画に投稿されたコメント約 300GB から成る [2]。

動画メタデータは、動画 ID、スレッド ID、動画タイトル、動画説明文、サムネイル画像の URL、投稿日時、再生数、動画フォーマット、高画質用ファイルサイズ、低画質用ファイルサイズ、再生数、コメント数、マイリスト数、

<sup>1</sup> 金沢工業大学基礎教育部  
Academic Foundations Programs, Kanazawa Institute of Technology

<sup>2</sup> 国立情報学研究所情報学ブリンシプル研究系  
Principles of Informatics Research Division, National Institute of Informatics

<sup>a)</sup> kurose@neptune.kanazawa-it.ac.jp

<sup>\*1</sup> <http://www.nicovideo.jp>

最近のコメント, およびタグ情報により構成される.

スレッド ID は動画のコメント情報取得の際に用いられる. マイリストは視聴者が気に入った動画を登録する機能で, Web や SNS (Social Networking Service) でのお気に入り登録やブックマークと同様の機能である. 各動画の再生数, コメント数, およびマイリスト数は視聴者から参照でき, 人気の指標として利用される.

ニコニコ動画では, 検索・分類のために動画に語句を付与することができる. これをタグと呼び, 1つの動画に複数付与できる. タグの追加, 編集, 削除は, コンテンツ投稿者または複数の視聴者により動的に行うことが可能である.

データセットのタグ情報は, 動画メタ情報の一部として保持されている. 1つの動画のタグ情報は, タグ文字列, カテゴリ, およびロックのタプルが複数含まれる. カテゴリ値が1の場合, そのタグ文字列は広く視聴者に利用される動画ジャンルと同様に扱うことができる. ロック値が1の場合, そのタグの編集・削除は投稿者以外には行うことができない.

他に動画に対するコメント情報があるが本稿では割愛する.

### 3. 解析

対象動画数は 8305696 で, 投稿期間は日本時間の 2007 年 03 月 06 日 00 時 33 分 00 秒から 2012 年 11 月 01 日 07 時 09 分 22 秒までの 2068 日の期間である. 動画の投稿・削除, タグの追加・編集・削除, 再生, コメント, マイリストは動的に行われている. データセットは 2012 年 11 月 1 日のスナップショットであるため, 本稿の解析結果は恒久的なものではない. 集計には, python 言語のライブラリ `numpy`<sup>\*2</sup>, `scipy`<sup>\*3</sup>を用いた.

動画の統計解析を 3.1 節に, タグの統計解析を 3.2 節に, 時系列データ解析を 3.3 節に, 動画メタ情報の出現頻度分布を 3.4 節に, タグ情報の出現頻度分布を 3.5 節に, タグの寿命をもとにした解析を 3.6 節に記す.

#### 3.1 動画統計解析

表 1 に動画の統計情報を示す. 集計はデータセットの全動画を対象としている. 列は左から項目 (item), 集計対象数 (samples), 平均値 (mean), 標準偏差 (stdev), 最小値 (min), 最大値 (max), 中央値 (median), 最頻値 (mode), 尖度 (kurtosis), 歪度 (skewness) の順である. 尖度は分布の変化の程度の指標で正規分布の場合に 0, 歪度は分布の非対称性の指標で左右対称の場合に 0 となる.

高解像度用ファイルサイズ file size (high) と低解像度用ファイルサイズ file size (low) の基数は 10 を用いている.

すなわち 1MB は 1000kB である. 高解像度用データレート data rate (high) は 1 秒に転送すべき最低データ量を意味し, この値を下回ると再生が中断する可能性がある. 低解像度用データレート data rate (low) も同様である. データレートはファイルサイズを再生長 length で除して求めている. ファイルサイズとデータレートの再頻度値は, ファイルサイズは MB 単位で, データレートは kB/s 単位で, 四捨五入している.

タイトル文字長 (title length), 動画説明文文字長 (description length), およびタグ長 (tag length) は UTF8 コードの文字数であり, 全角でも 1 文字である. タグ長は, タグの種類別に求めた distinct tag と動画に付与された attached tag の 2 種があり, 後者は同一タグが集計対象に複数含まれている. これらは, タグ情報であるがタイトルおよび動画説明文と比較するためここに掲示する.

動画フォーマットの内訳は, mp4 (MPEG-4) が 5553371 動画, swf (Small Web Format) が 354515 動画, flv (Flash Video) が 2397810 動画であり比は, 67:4:29 である.

#### 3.2 タグ統計解析

表 2 にタグの利用状況を示す. 集計項目は, 3.1 節と同様である. 1つの動画には, 複数のタグが付与できるため動画に割り当てられたタグ数は動画数を上回る. ここで, distinct tag は動画に付与されたタグの種類を表し, 5328340 種類あり, 平均 8.95 回動画に付与され, 最大 3593302 動画に付与されたタグがある.

category tag は, distinct tag のうち, カテゴリが指定されたタグ (タグ情報の category 値が 1) である. 同様に locked tag はロックが指定されたタグ (タグ情報の lock 値が 1) を, lock&category は双方が指定されたタグである. タグ情報は編集が可能であり, 動画情報内に保存されるため, 同一タグであっても category または lock の値が異なるものが存在する. それらの内訳は表 3 で後述する.

タグでは, `&`, `>`, `'`, `<` が HTML と同様に `&amp;`, `&gt;`, `&quot;`, `&lt;` で格納されており, それぞれ 5341, 2361, 1604, 1690 種のタグで使用されていた. これらのエスケープ文字を持つタグは 10301 種であった.

#### 3.3 時系列解析

図 1 に日別動画投稿数を, 図 2 に日別投稿データ量を示す. 開始日は 2007 年 3 月 7 日からである. 最終日は 7 時 9 分 22 秒までのため投稿が少なく図から除外している. 変化の傾向を見るため実線で 7 日間の移動平均を示す. いずれも日数が経過するにつれて増加している.

最小二乗法による 1 次式近似で, 経過日数  $days$ , 投稿数  $nPosted$ , 投稿量  $sPosted$  [TB] で表すと式 (1), (2) となる.

\*2 [www.numpy.org](http://www.numpy.org)

\*3 [www.scipy.org](http://www.scipy.org)

表 1 動画メタデータの統計 (動画数=8,305,696)

Table 1 statistics of video meta data(number of videos is 8,305,696)

item	samples	mean	stdev	min	max	median	mode	kurtosis	skewness
length [sec]	8305518	666.705	754.739	1	65535	388	90	447.32	8.37
number of video posted/day	2068	4016.294	1613.387	310	8381	4240	4014	-0.46	-0.32
amount of posted [TB/day]	2068	139.491	86.307	3	371	137	60	-1.04	0.20
number of views	8305696	4174.734	31870.100	0	15454295	529	24	19398.77	82.97
number of comments	8305696	297.764	14067.464	0	22830467	20	0	1745705.59	1166.95
number of mylists	8305696	67.764	805.736	0	517528	3	1	30960.23	102.20
tags per content	8305696	5.511	2.735	0	12	5	4	-0.84	0.15
title length	8305696	25.693	10.089	1	212	26	27	1.32	0.38
description length	8305696	164.809	125.008	0	1543	140	15	0.76	0.93
distinct tag length	5328340	8.831	4.192	1	40	8	6	1.85	1.05
attached tag length	45773378	6.198	3.681	1	40	5	3	3.13	1.43
file size (high) [MB]	8305459	35.572	32.360	8.774e-05	503	26	38	-0.08	0.90
file size (low) [MB]	8305459	21.400	20.921	3.719e-05	825	14	3	6.66	1.76
data rate (high) [kB/s]	8305281	73.826	98.551	3.740e-06	40603	52	38	5565.89	24.78
datarate (low) [kB/s]	6779295	36.398	13.618	2.529e-05	10762	38	38	56780.47	71.70

表 2 タグ利用状況

Table 2 Tag usage situation(number of tags is 5,328,340)

item	samples	mean	stdev	min	max	median	mode	kurtosis	skewness
distinct tag	5328340	8.591	1748.396	1	3593302	1	1	3396621.39	1718.13
category tag	103	66371.466	342091.867	1	3365363	12	1	81.17	8.81
locked tag	816193	24.068	4083.702	1	3406341	1	1	596455.40	731.52
lock&category	103	66369.311	342091.274	1	3365363	12	1	81.17	8.81

$$nPosted = 2.455days + 1482 \quad (1)$$

$$sPosted = 0.150days - 5.354 \quad (2)$$

移動平均および1次式近似のy切片から、投稿数の増加は投稿量の増加より伸びが鈍化している。これは、投稿される動画数は飽和傾向にあるが、動画あたりのファイルサイズが大きくなっているためと推定する。ここで、留意事項として投稿された動画は投稿者または運営により削除される場合があり、データセット内の動画は削除されたものが含まれているか除外されているか判断不能である。

### 3.4 動画メタ情報の出現頻度分布

以下図中のbinsは区間分割数である。

図3に日別動画投稿数の累積分布を示す。1日あたり3000から6000動画が投稿される確率が同程度存在する。1日に4000動画以内が投稿される確率が約50%である。

図4に日別投稿量の累積分布を示す。1日あたりの投稿量が250TBまでの確率が同程度存在する。

図5に動画再生時間の累積分布を示す。動画メタ情報の再生長lengthを分単位で表示している。再生時間が0秒のデータが178件については集計から除外した。動画の再生時間は平均11分程度である。30分以内の動画が全体の

90%を占めている。

図6に動画ファイルサイズの累積分布を示す。低解像度用ファイルサイズが0で高解像度用ファイルサイズが設定されている場合が1526270件あり、この場合は、高解像度のファイルサイズを採用した。高解像度、低解像度ともにファイルサイズが0の場合が237件あり、これらは集計から除外した。高解像度と低解像度とも生起確率は同様に、10から100MBの動画が全体の90%以上を占める。

図7にデータ転送レートの累積分布を示す。このレートは、1秒間に転送するデータ量の下限であり、このレートを下回ると視聴が中断する可能性がある。レートは、データセット中のファイルサイズを再生長で除して求めている。除外するデータの条件は、図5、図6と同様である。大多数の動画は網羅するには、低解像度では55KB/s付近、高解像度では、1MB/s付近の転送レートが必要である。

図8に再生数、コメント数、マイリスト数の分布を示す。スケールフリー[3]の傾向を見るため両軸とも対数として表示している。スケールフリーは、両対数グラフ上の頻度分布が負の傾きを持つ直線となる特長を持つ。コメント数、マイリスト数は、ほぼ直線的に分布した。再生数では、再生が少ない動画の頻度が少ないため、コメントの分布と交差している。これは、投稿されたばかりの動画や視聴の機会が少

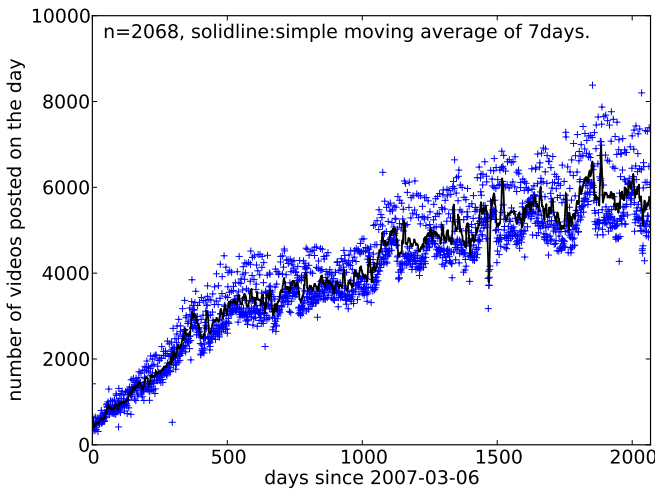


図 1 日別動画投稿数

Fig. 1 number of video posted on each day

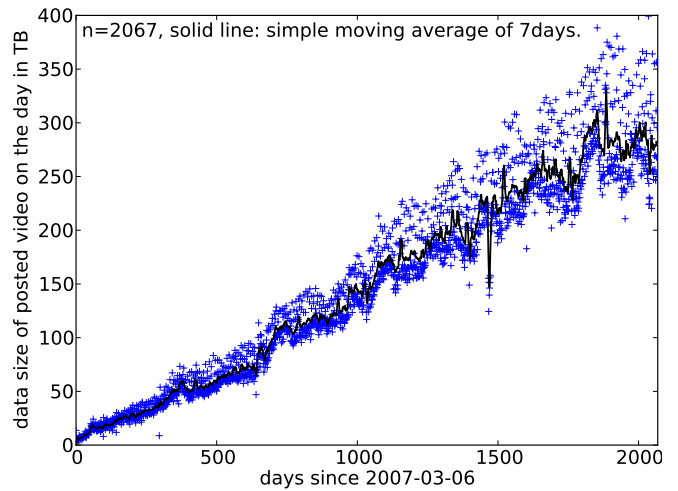


図 2 日別投稿動画量

Fig. 2 total size of posted files on each day

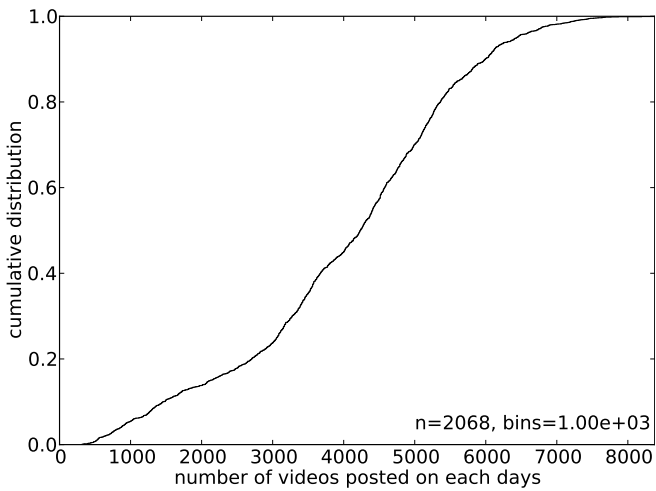


図 3 日別動画投稿数の累積確率分布

Fig. 3 cumulative distribution of posted videos number

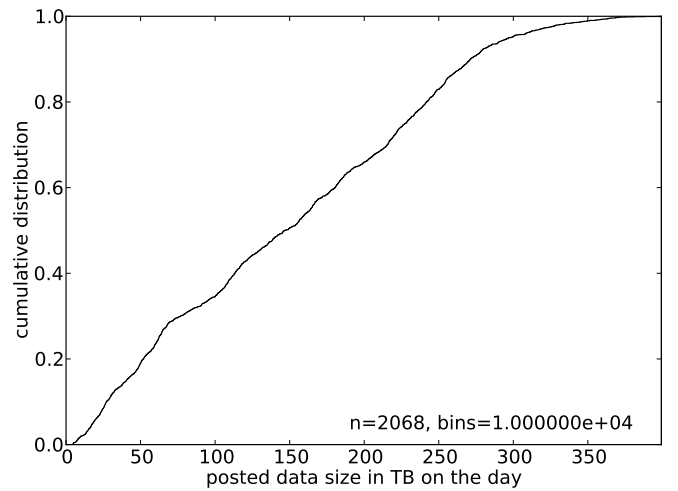


図 4 日別投稿量の累積確率分布

Fig. 4 cumulative distribution of posited videos size

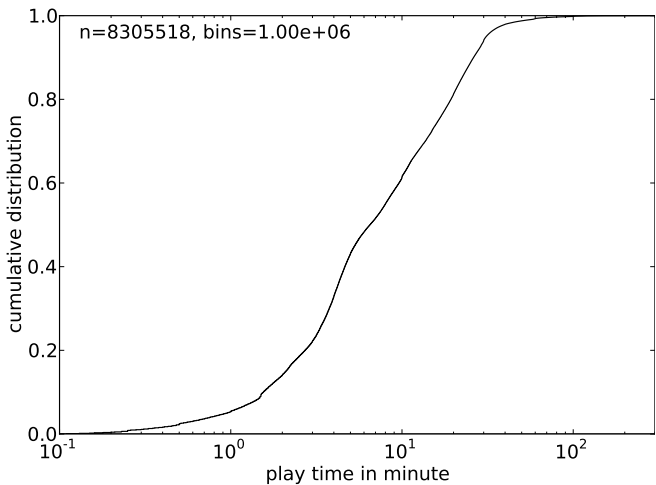


図 5 動画再生時間の累積確率分布

Fig. 5 cumulative distribution of play time

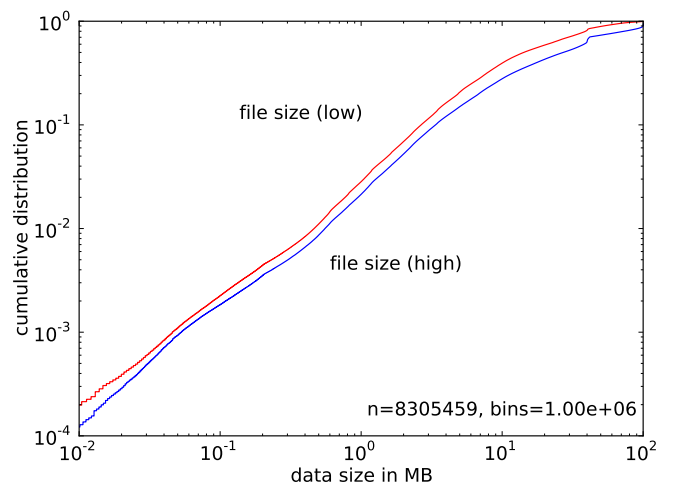


図 6 動画ファイルサイズの累積確率分布

Fig. 6 cumulative distribution of video file size

ない動画が数多くあるためと考えられる。コメント数、マイリスト数は、ほぼ直線的に分布した。

図 9 に動画タイトル、動画説明文、タグの文字長の出現分布を示す。ここで文字長は UTF8 コードの文字長で

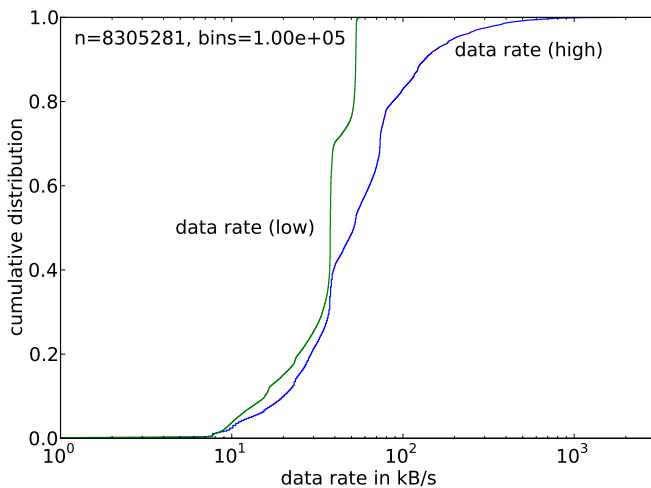


図 7 データ転送レートの累積確率分布  
**Fig. 7** cumulative distribution of data rete

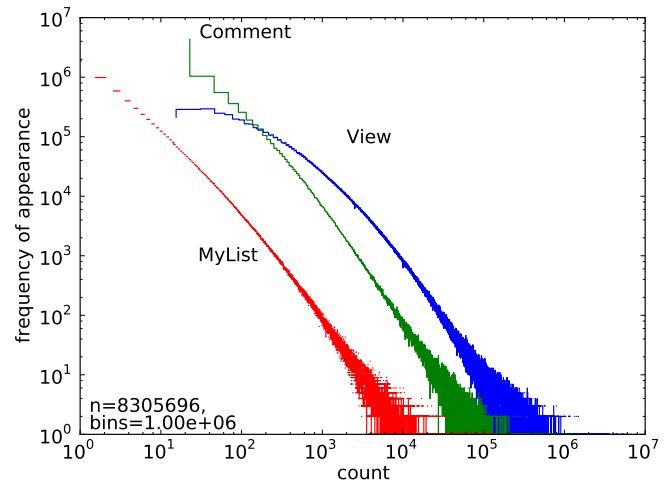


図 8 再生数, コメント数, マイリスト数の出現頻度分布  
**Fig. 8** appearance of counts of views, comments, mylists

ある。タグはタグ種別 (distinct tag) と動画に付与されたタグ (attached tag) の2つを集計している。タイトルは30字程度が、動画説明文は40字以内が多い。タグは、種別では平均8.3字、動画付与では平均6.1字になっていることから、語句の文字長が短いタグが動画に多く付与されている。タグ付与第1位は「ゲーム」でこの場合は3字である。

### 3.5 タグ情報の出現頻度分布

図10に動画に付与されたタグ数の頻度分布を示す。動画あたりのタグ数は平均5.51であった。タグが付与されない動画は全体の2%存在した。頻度では、1動画に付与されたタグ数は4か5が多いが両方合わせても全体の約30%である。

図11にタグが付与された動画の多い順から並べたタグランキングの累積分布を示す。カテゴリタグは種類も少なく第1位で60%を占め上位20位まででカテゴリタグが付与された動画の大多数を網羅する。全タグを対象とした場合は、第10<sup>4</sup>位を超えないとタグが付与された動画の過半数を網羅できない。

最小二乗法による指数近似  $a^{bx}$  では、ランキング順位  $k|1 \leq k$  に対してタグ、ロック指定タグ、カテゴリ指定タグの出現数は、それぞれ式(3)、(4)、(5)である。

$$Tag = 3.723 \exp(-3.56e-7 \cdot k) \quad (3)$$

$$Tag_{lock} = 13.510 \exp(-4.349e-6 \cdot k) \quad (4)$$

$$Tag_{category} = 1.552e+5 \exp(-0.1475 \cdot k) \quad (5)$$

同一タグであってもカテゴリまたはロックが指定されている場合(値が1)と指定されていない場合があるため、状況について確認する。カテゴリ、ロックの値を変えてタグ数を調べた結果を表3に示す。anyは値が任意であることを表す。number of attachedは動画に付与されたタグで、

distinct tagは種別である。

表 3 タグ割当状況 (動画数=8,305,696)  
**Table 3** Attached tag(# of videos is 8,305,696)

category	lock	number of attached	distinct tag
0	0	26,128,931	4,939,008
0	1	12,808,186	816,157
0	any	38,937,117	5,328,312
1	0	222	10
1	1	6,836,039	103
1	any	6,836,261	103
any	0	5,328,312	4,939,008
any	1	19,644,225	816,193
any	any	45,773,378	5,328,340

図12にカテゴリ指定されたタグの動画への付与数を示す。横軸は、付与数の多いタグ順に並べている。動画への付与数は、30位を超えると急速に減少する。

カテゴリ指定のタグは103種類あるが、タグの語句を見ると、「VOCALOID」と「vocaloid」のように表記ゆれを同一とすると41種となる。さらに「日記」と「tagebuch」など意味が同じタグを同一視すると33種に集約できる。このうち動画に付与された数が2以上のものは30種であった。なお、ニコニコ動画のタグ検索機能では、検索文字の大文字・小文字を同一視している。

次に、あるタグが付与された動画の総再生数とタグの出現数の関係を見る。図13にタグによる再生数ランキング分布を示す。タグ出現数のランキングを上位100位までの各タグにおいて、そのタグが付与された動画の総再生数を求め、再生数の多い順に並べ直している。

同様にタグ出現順位を100位ずつ増やして800位までの分布を描く。第100位までの最低再生総数は10<sup>7</sup>程度であるが、第200位での最低再生総数は、5×10<sup>5</sup>程度になる。これは、動画に付与される数が多いタグでも再生数が低い

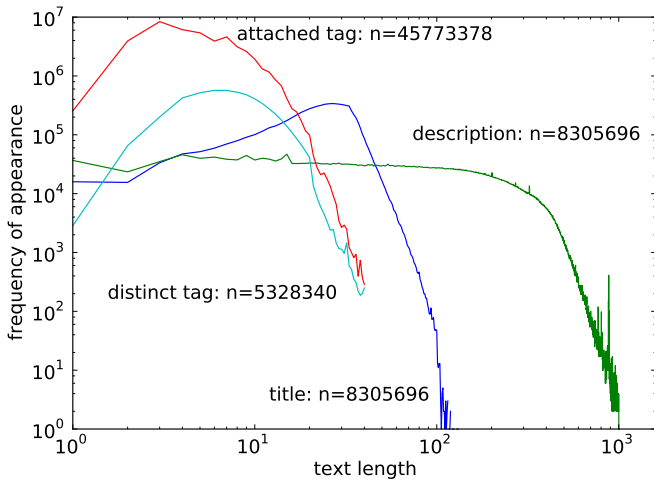


図 9 文字長の出現頻度分布  
Fig. 9 appearance of text length

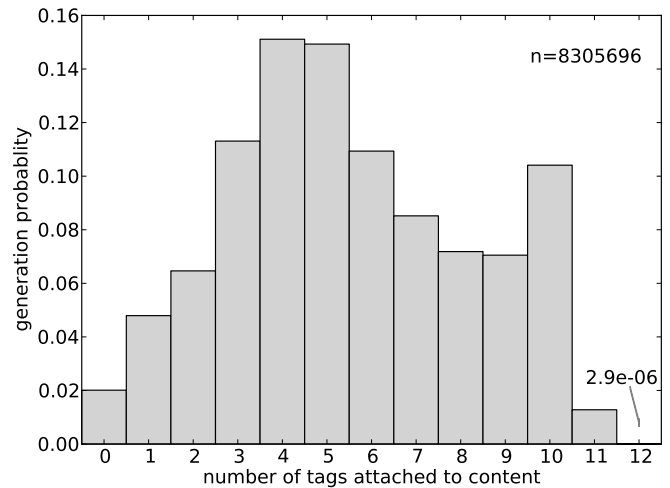


図 10 タグ割当数の出現確率  
Fig. 10 number of tags attached to a content

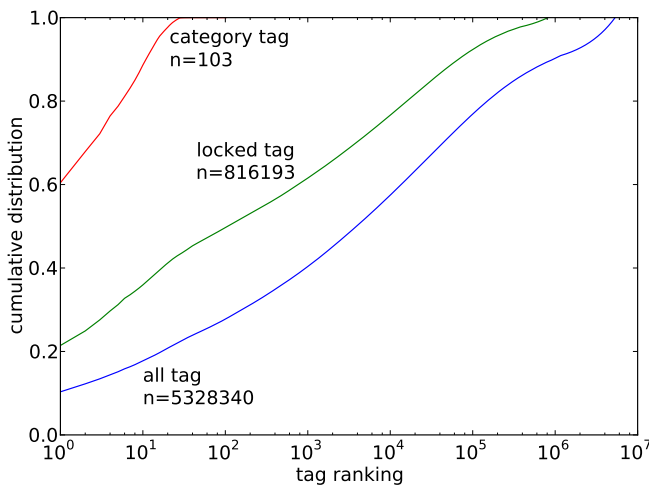


図 11 タグ付与の累積分布  
Fig. 11 cumulative distribution of attached tag

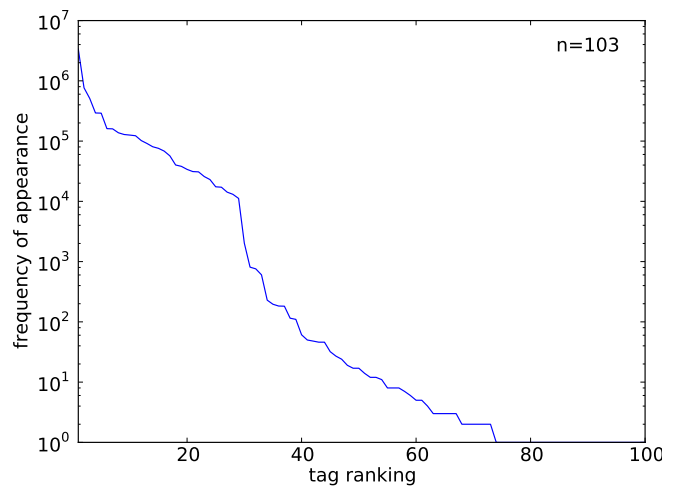


図 12 カテゴリタグの付与数  
Fig. 12 number of videos using category tag

ものが存在していることによる。上位 600 位までは、上位 200 位までの最低再生数を下回るタグは出ていない。

タグ出現ランキング  $n$  位までの再生数が、タグによる再生総数にしめる割合により正規化する。動画の再生数  $nViews$ 、動画総数  $nV$ 、動画に付与されたタグ数  $nTags$ 、動画に付与されたタグ集合  $tag_j$  とすると式 (6) となる。これは、タグ出現頻度  $n$  位までのタグによる再生総数が、タグによる再生機会にしめる割合すなわち網羅率  $C_{n,view}$  である。

$$C_{n,view} = \frac{\sum_{i=1}^n \sum_{j=1}^{nV} nViews_{j,tag_i \in tag_j}}{\sum_{i=1}^{nV} nViews_i \cdot nTags_i} \quad (6)$$

同様にコメント網羅率  $C_{n,comment}$ 、マイリスト網羅率  $C_{n,mylist}$  を導入する。

図 13 に網羅率を付記する。100 位までで全体の 1.70%、800 位までで 2.77% の再生数を網羅する。

図 14 にタグ出現順位によるコメント総数を図 15 にタグ出現順位によるマイリスト総数を示す。これらは、

図 13 と同様の傾向が見られた。view の最低値は、上位 100 位で 12734554 が、上位 200 位で 562206 に、上位 700 位で 385662 になった。comment の最低値は、上位 100 位で 961382 が、上位 200 位で 17465 に、上位 700 位で 5450 になった。mylist の最低値は、上位 100 位で 108621 が、上位 200 位で 13346 に、上位 700 位で 6727 になった。いずれも同様の傾向が得られた。

最後に正規化した再生数、コメント数、マイリスト数の網羅率を図 16 に示す。いずれの網羅性も同様の値を得た。上位 10 位までのタグにより 10% の網羅性があるが、30% を網羅するには上位 1000 位までが必要となる。網羅率の最小二乗法による指数近似を式 (7), (8), (9) に示す。

$$C_{2 \times 10^5, view} = 0.1883 \exp(1.248e-5 \cdot k) \quad (7)$$

$$C_{2 \times 10^5, comment} = 0.1891 \exp(1.183e-5 \cdot k) \quad (8)$$

$$C_{2 \times 10^5, mylist} = 0.1726 \exp(1.376e-5 \cdot k) \quad (9)$$

最初は、マイリスト網羅率が低くコメント網羅率が高いが、



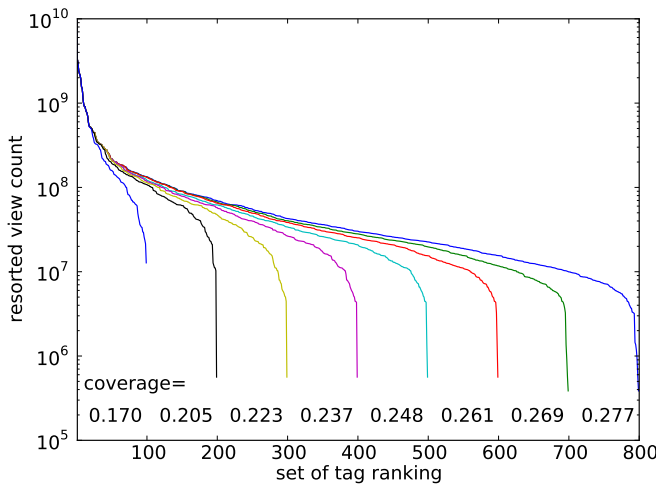


図 13 タグランキングによる総再生数  
Fig. 13 total view count by tag ranking

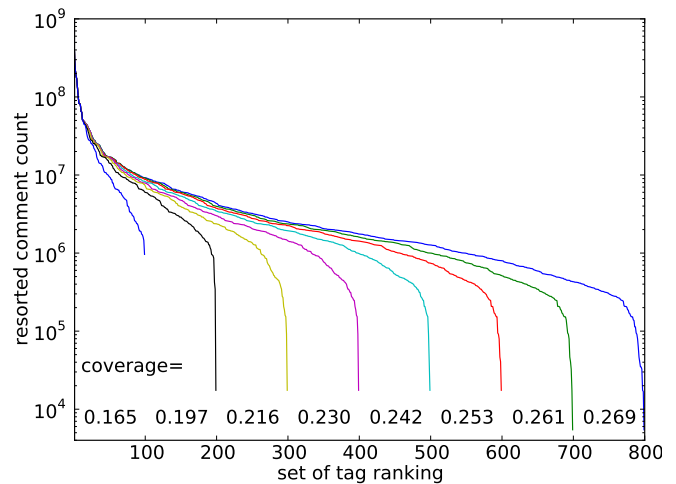


図 14 タグランキングによる総コメント数  
Fig. 14 total comment count by tag ranking

1000 以降では逆転するが値に大きな変化は見られない。タグランキングが全体の 38%程度である  $2 \times 10^5$  位でもタグ付与動画数は 16 であり網羅率は約 70%である。

### 3.6 タグの寿命

タグが利用される期間に着目して少ないメモリ容量で検索の網羅性を向上することが可能か確認する。

実際には、タグの付与・削除は動的に行われているが、動画メタデータには、タグ編集履歴が含まれないため、動画投稿時にタグが付与されたものとして扱うことにする。

あるタグが付与された動画のうち、日付が一番古いものと一番新しいものの投稿日の差を寿命とする。タグの寿命とそのタグが付与された動画数の関係を図 17 に示す。ここでは、10000 位までのタグランキングを順位により区分して傾向を見る。図の右へ行くほど寿命が長いタグを表す。10 位以内は集計期間の最初から最後まで動画に付与されている。ただし 2 位の「実況プレイ動画」の寿命は 2058 日である。いずれのタグもデータ収集期間のほとんどで利用されている。11 から 100 位までは、2 年以上の間使用されている。101 から 1000 位までは、寿命が広範囲となり、数日で 2 万再生された動画も含まれる。以下、順位が下がるほど、広範囲の寿命を持つ。縦軸は、タグが付与された動画数であるため、この図では、タグランキングの順位により層をなす。ここで、タグは、明示的に使用禁止とならないため、データセットの収集期間を長くすると右軸にシフトしていくと考える。

次に、縦軸をタグによる総視聴数をそのタグが付与された動画数で除した平均視聴数にしてタグの寿命との関係を図 18 に示す。この散布図は、図 17 に比べ縦軸の値に差が大きいため、タグランキング 1000 位までをプロットする。第 10 位までは、図 17 と同傾向であるが、11-100 位、101-200 位でも平均再生数が大きなタグが出現している。

平均再生 10000 回以上のタグでは、ランキング 700 位以内かつ寿命 1 年以上の範囲に入っている。

タグの寿命と平均再生数から、定常的に利用されるタグや急速に利用が進んでいるタグを検出する方法を検討することが今後の課題である。

## 4. コンテンツ情報流通基盤への影響

フォークソノミーでは、タグによる参照数が多い順にランキングすると参照数はロングテール状になることが知られている [4]。コンテンツを中心とした配信基盤では、少数の参照だが数が多いタグの扱いにより検索精度や性能に影響が大きいため、方式の検討が行われている。集中管理での評価 [5] や分散管理での評価 [6] がある。Web 検索では検索精度を高めると網羅性が下がるが、タグを用いた場合も同様の傾向がある [7]。フォークソノミータグをキーとしたルーティングでは、コンテンツに割り当てるタグ数やキャッシュ容量が検索およびダウンロード性能に影響を与える [8]。

タグをキーとしてコンテンツ探索をネットワーク上の複数のルータが分散処理で行う情報流通基盤を考える場合に 3 節の結果より以下のことが利用できる。

- ニコニコデータセットでは再生数、コメント数、マイリスト数いずれを用いても同様の網羅率となる
- タグランキングのロングテール部分が大きいため網羅率の収束性はよくない
- タグは利用され続けるためタグランキングでは、上位タグの変動が少ないと考えられる
- カテゴリタグは固定的に使用され数が少ないので動画に付与されたカテゴリタグと別のタグの組み合わせにより網羅性を改善する方法が求められる
- タグ検索を行う場合にタグランキング後半以降の寿命と最終利用日を用いる方法が考えられる

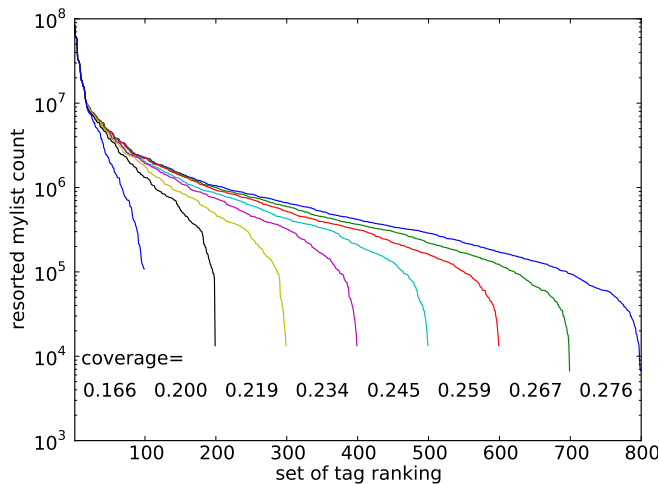


図 15 タグランキングによる総マイリスト数  
Fig. 15 total mylist count by tag ranking

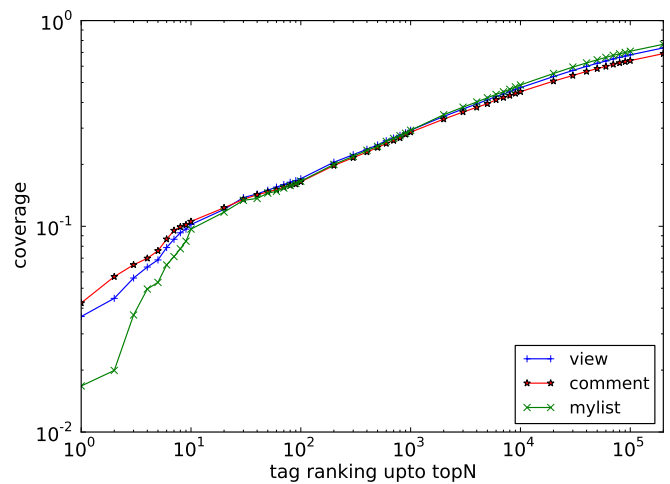


図 16 タグランキングによるカウント値の網羅性  
Fig. 16 coverage the counts by tag ranking

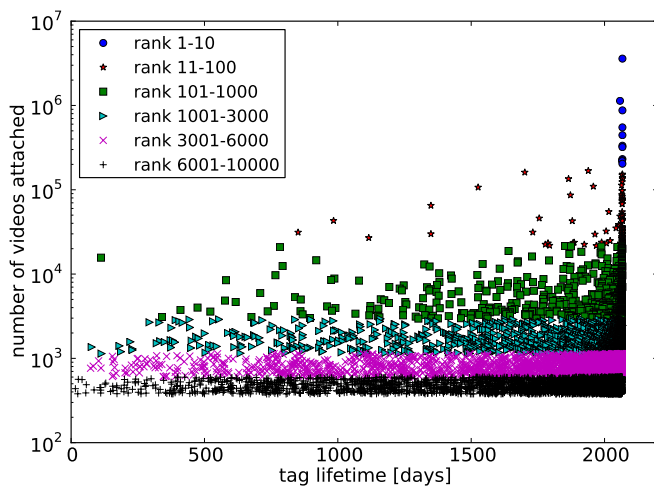


図 17 タグの寿命と付与数  
Fig. 17 tags duration vs. number of attached

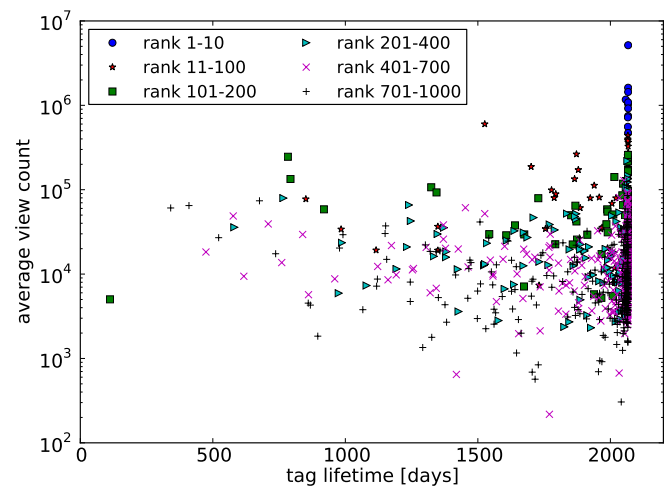


図 18 タグの寿命と平均再生数  
Fig. 18 tags duration vs. average view counts

## 5. おわりに

スナップショットであるニコニコデータセットを解析し、コメント数、参照数、マイリスト数いずれも同様のタグランキングの傾向があることを確認した。タグによる検索では、タグ名のゆらぎを集約する機構やランキング後半のコンテンツ割当数の少ないタグの検索効率を高める機構が求められる。

謝辞 ニコニコデータセットの解析にあたり情報学研究所リポジトリを使用した。株式会社ドワンゴおよび国立情報学研究所に謝意を表す。

### 参考文献

[1] 国立情報学研究所：情報学研究データリポジトリ, <http://www.nii.ac.jp/cscenter/idr/>.  
[2] 国立情報学研究所：ニコニコデータセット, <http://www.nii.ac.jp/cscenter/idr/nico/nico.html>.  
[3] Caldarelli, G., Capocci, A., De Los Rios, P. and Muñoz, M. A.: Scale-Free Networks from Varying Vertex Intrinsic Fitness, *Phys. Rev. Lett.*, Vol. 89, p. 258702 (online),

DOI: 10.1103/PhysRevLett.89.258702 (2002).

[4] Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G.: FolkRank: A ranking algorithm for folksonomies, *UNIVERSITY OF HILDESHEIM, INSTITUTE OF COMPUTER SCIENCE*, pp. 111–114 (2006).  
[5] BibSonomy: BibSonomy - The blue social bookmark and publication sharing system, <http://www.bibsonomy.org/>.  
[6] Rossini, G. and Rossi, D.: Large scale simulation of CCN networks, *14èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, La Grande Motte : France (2012)*, (online), available from (<http://hal.archives-ouvertes.fr/hal-00688934/>) (2012).  
[7] Wetzker, R., Zimmermann, C., Bauchhage, C. and Albayrak, S.: I tag, you tag: translating tags for advanced user models, *Proceedings of the third ACM international conference on Web search and data mining, WSDM '10*, New York, NY, USA, ACM, pp. 71–80 (online), DOI: 10.1145/1718487.1718497 (2010).  
[8] 黒瀬 浩, 山田茂樹, クリスチャンボルセア: TCR:フォークソノミータグに適したCGMコンテンツキャッシュルーアーキテクチャ, 電子情報通信学会論文誌. B, 通信, Vol. 96-B, No. 2, pp. 71–82 (2013).