

YouTube チャンネルの投稿動画数や動画再生回数の推移に基づく登録者数分析

Analysis on Subscribers of YouTube Based on Changes

in the Number of Posted Videos and Video Views

松清 綾大^{†1} 伊藤 淳子^{†2} 吉野 孝^{†2} 宗森 純^{†3}
 Ryota Matsukiyo Junko Itou Takashi Yoshino Jun Munemori

1. はじめに

YouTube[1]上に自身の動画配信チャンネルを開設し、オリジナルの動画や音楽コンテンツを制作しアップロードする YouTuber と呼ばれる個人や団体が多く活動している。YouTube にはチャンネル登録という機能があり、視聴者は好きな YouTube チャンネルを登録するとそのチャンネルに関する通知などを受け取ることができる。1つのチャンネルに対し何人がチャンネル登録をしているかを表すチャンネル登録者数は、YouTuber の人気や認知度を示す指標としてたびたび用いられる。また、人気があり、チャンネル登録者数が多い YouTuber は視聴者の生活習慣や消費行動に大きな影響力を持つため、YouTuber を起用したマーケティングを行う企業なども存在し、その数は増加している [2]。このことから、チャンネル登録者数の分析や予測は様々な面で有用であると考えられる。

これらの背景を踏まえて、YouTube Data API[3]から取得したチャンネル登録者数、投稿動画数、チャンネル全体の動画再生回数(以下、動画再生回数という)を使用し、1か月後のチャンネル登録者数と1か月間のチャンネル登録者増加数を予測する線形回帰モデルの作成を試みた [4]。その結果、過去1週間でチャンネル登録者数が増加したチャンネルは、その後1か月はチャンネル登録者数が増加する傾向にあると分かった。しかし、特に登録者数が増加した YouTube チャンネルが持つ傾向などについては明らかにならなかった。そこで本稿では、分析対象を新たに開設された YouTube チャンネルに限定し、チャンネル登録者数、投稿動画数、動画再生回数の推移データを取得する。これらのデータを用いて、特に登録者数が増加した YouTube チャンネルがどのような傾向にあるか分析をする。また、最も単純な学習アルゴリズムと言われている k-最近傍法を利用し、特にチャンネル登録者数が増加した YouTube チャンネルを予測する分類器の作成と精度の評価を行い、チャンネル登録者数の推移を分析する。

2. 関連研究

田中らは、YouTube 上で高い人気が続く動画を単純ベイジ分類器で判別する場合の予測精度の評価を行った [5]。その結果、アップロードから3時間経過した視聴数推移データを用いた場合、初期の視聴数の絶対値のみで予測する場合よりも高精度で予測できることを明らかにした。このような、推移データを用いた分類器の作成はチャンネル登録

表 1: 取得可能なチャンネル登録者数の省略方法

チャンネル登録者数 (人)	チャンネル登録者数が更新される単位 (人)
~1,000	1
1,000~9,999	10
10,000~99,999	100
100,000~999,999	1,000
1,000,000~9,999,999	10,000
10,000,000~99,999,999	100,000
100,000,000~999,999,999	1,000,000

者数が増加した YouTube チャンネルの予測にも有効な可能性があると考えられる。

3. データの取得と取得データの分析

3.1 データの取得方法

新たに開設された YouTube チャンネルの登録者数分析を行うために、YouTube Data API を使用し、チャンネル開設から1日ごとの推移データを取得した。

チャンネル開設からの推移データを取得するために、2021年2月22日~2021年6月8日までの間、YouTube Data API からデータを取得する下記の2つのプログラムを毎日12時に実行した。

- 前日に開設されたチャンネルのチャンネル ID を取得
- 取得したチャンネル ID に該当するチャンネルの登録者数、投稿動画数、動画再生回数を取得

この取得方法で 23,556 チャンネルの登録者数、投稿動画数、動画再生回数の 60 日間の推移データを取得した。この 23,556 チャンネルを分析対象チャンネル A とする。

YouTube Data API で取得可能なチャンネル登録者数はその数に応じて省略されている [6]。例えば、チャンネル登録者数が 1,000~9,999 人の場合、10 人単位の増減でチャンネル登録者数が更新され、1 の位が省略された値が取得可能である。具体的な省略方法は表 1 に示す。

3.2 チャンネル登録者数の分布

前述の分析対象チャンネル A について、チャンネル開設から 60 日後の登録者数の分布を表 2 に示す。表 2 から分かるように、チャンネル開設から 60 日後において、分析対象チャンネル A のうち 52.7% は登録者数が 0 人であり、登録者数が 1,000 人以上存在する YouTube チャンネルは、43 チャンネルであった。また、チャンネル登録者数が 1,000 人以上であることが、YouTube パートナープログラム [7] の利

^{†1} 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

^{†2} 和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

^{†3} 放送大学 和歌山学習センター, Wakayama Study Center, The Open University of Japan

表 2: 分析対象チャンネル A の桁数に応じた分布

チャンネル登録者数	チャンネル数 (開設から 60 日後)
0 人	12,422
1 人以上,10 未満	8,534
10 人以上,100 人未満	2,302
100 人以上,1,000 人未満	255
1,000 人以上,10,000 人未満	39
10,000 人以上,100,000 人未満	4

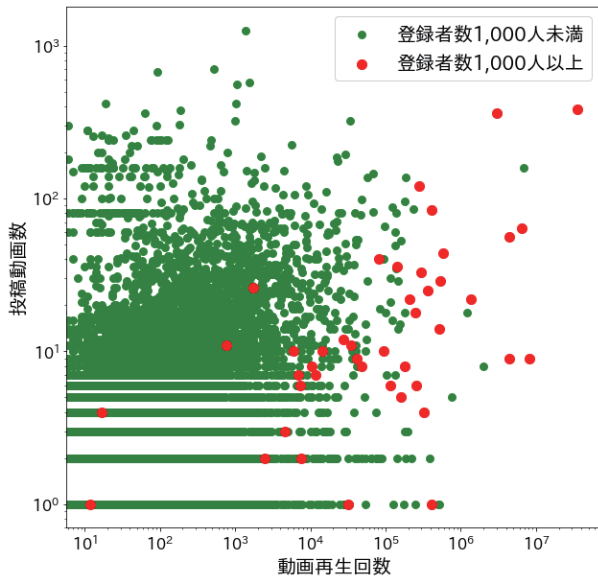


図 1: チャンネル開設から 60 日後のデータの散布図

用資格の最小要件の一つであることから、本研究では、チャンネル開設から 60 日後に登録者数が 1,000 人以上存在する YouTube チャンネルを「特に登録者数が増加した YouTube チャンネル」として扱う。

3.3 チャンネル登録者数と動画再生回数の関係

分析対象チャンネル A のチャンネル開設から 60 日後のデータについて、横軸が動画再生回数、縦軸が投稿動画数の両対数グラフでの散布図を作成した。その結果を図 1 に示す。図 1 において、特に登録者数が増加した YouTube チャンネルが右側に多く見られ、それ以外のチャンネルが左側に多く見られることから、特に登録者数が増加したチャンネルは比較的動画再生回数が多い傾向にあると考えられる。

4. 分類器の作成と精度の評価

4.1 分類器の作成方法

特に登録者数が増加した YouTube チャンネルを予測する分類器を作成するために、最も単純な学習アルゴリズムであると言われている k-最近傍法を利用する。なお、特に登録者数が増加した YouTube チャンネルのデータをクラス 1 のデータとして扱う。

チャンネル開設から $N(1 \leq N \leq 59)$ 日後までの推移データの特徴量として使用し、クラス 1 かそれ以外かを分類する分類器を作成する。分類器の作成には Python の機械学習

ライブラリ scikit-learn[8] を使用する。また、特徴量としてチャンネル開設から N 日後までの以下のデータを使用した 4 つの分類器をそれぞれ作成し、評価を行う。

- チャンネル登録者数の推移データ
- 動画再生回数の推移データ
- 投稿動画数の推移データ
- チャンネル登録者数、動画再生回数、投稿動画数の 3 つ全ての推移データ

4.2 分類器の評価方法

分類器の評価は層化 3 分割交差検証による適合率 (Precision)、再現率 (Recall)、F 値 (F-measure) の 3 つのそれぞれの平均値により行う。適合率、再現率、F 値はそれぞれ式 (1)、(2)、(3) と定義する。なお、TP, FP, FN, TN は以下のように定義する。

- 真陽性 (TP)：正解がクラス 1 のデータを正しくクラス 1 に分類できたデータ数
- 偽陽性 (FP)：正解がクラス 1 以外のデータを誤ってクラス 1 に分類したデータ数
- 偽陰性 (FN)：正解がクラス 1 のデータを誤ってクラス 1 に分類したデータ数
- 真陰性 (TN)：正解がクラス 1 以外のデータを正しくクラス 1 以外に分類できたデータ数

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F\text{-measure} = \frac{2TP}{2TP + FP + FN} \quad (3)$$

近傍点の数 k を決定するために、チャンネル登録者数の推移データの特徴量とする $N=59$ の分類器を作成した。k=1 ~ 10 の時の層化 3 分割交差検証による適合率、再現率、F 値の 3 つのそれぞれの平均値を図 2 に示す。k=3 の時に適合率、再現率、F 値の平均値がそれぞれ大きいことから、本章では k=3 とし、分類器を作成する。

4.3 分類器の精度の検証

作成した分類器の精度を図 3 に示す。図 3 から分かるように、N が 2 以上のとき、チャンネル登録者数の推移データを使用したモデルが再現率、F 値において 4 つのモデルの中で最も精度が高いモデルとなった。適合率においても $N=23,24$ の場合を除き、同様の結果となった。また、チャンネル登録者数の推移データを使用したモデルは N が 42 以上のとき、適合率、再現率、F 値全てにおいて 0.8 以上の高い値をとった。

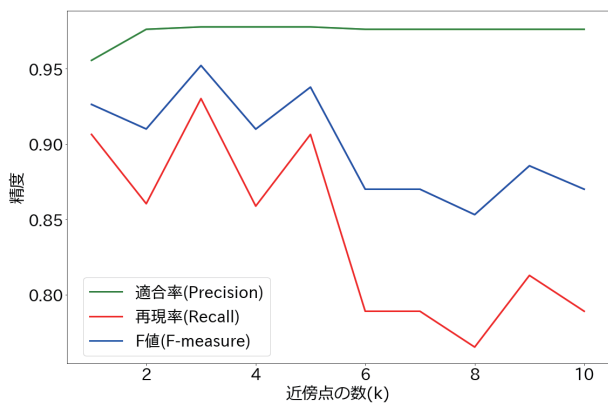


図 2: k=1~10 における k-最近傍法の精度

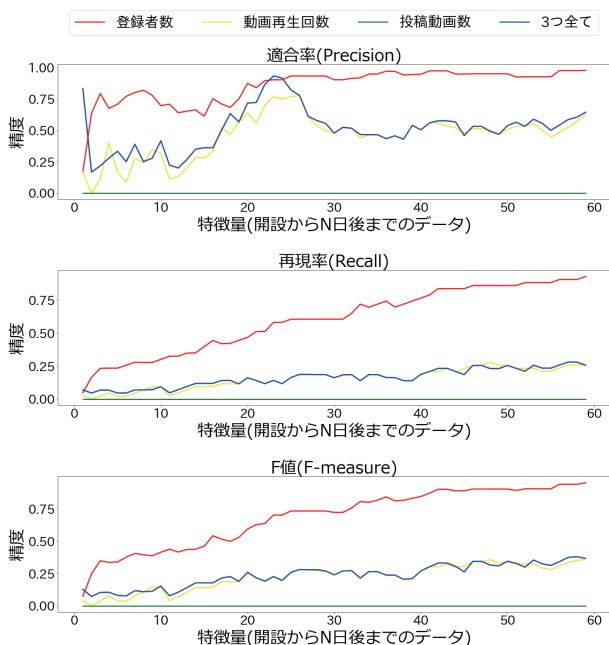


図 3: 分類器の精度

4.4 考察

4.3 節に示した結果から、チャンネル登録者数の推移データの特微量として予測を行う場合、より長期間の推移データを利用の方が予測の精度が上がるということが分かった。一方、投稿動画数の推移データの特微量として使用することは、より長期間の推移データを利用しても精度が上がらないことが明らかになった。このことから、投稿動画数の推移データはチャンネル登録者数に大きく影響しないと考えられる。また N=23,24 のとき、チャンネル登録者数、動画再生回数、投稿動画数の 3 つ全ての推移データを使用したモデルの適合率が高い値を取るが、再現率、F 値においてその傾向が見られないことから、このモデルが精度の高いモデルであるとは言い難い。

5. おわりに

本稿では、新たに開設された YouTube チャンネルの登録者数、投稿動画数、動画再生回数の推移データを用いて、特に登録者数が増加した YouTube チャンネルの傾向について

分析を行った。その結果、チャンネル開設から 60 日後において登録者数が 1,000 人を超える YouTube チャンネルはごく少数であり、それらのチャンネルは動画再生回数が多い傾向にあることが分かった。

また、k-最近傍法を使用し特にチャンネル登録者数が増加した YouTube チャンネルを予測する分類器の作成と精度を評価した。その結果、より長期間の推移データを利用の方が予測の精度が上がることや動画再生回数、投稿動画数の推移データの特微量とした分類器のモデルでは高い精度が得られないことが分かった。

今後の展望として、より長期間の推移データの特微量とする分類器の作成や、k-最近傍法以外のアルゴリズムを使用した分類器の精度の評価を検討する。

参考文献

- [1] YouTube ”<https://www.youtube.com/>” (2021.07.09 確認).
- [2] 鎌田和樹：講座開講記念講演「ネット動画マーケティングから見るクリエイター育成とビジネス開発の可能性」, DHU journal = デジタルハリウッド大学紀要 / デジタルハリウッド大学編, No.4,pp.33-38(2017).
- [3] YouTube Data API ”<https://developers.google.com/youtube/v3>” (2021.07.09 確認).
- [4] 松清綾大, 伊藤淳子, 宗森純：機械学習を用いた YouTube チャンネル登録者数の予測, 2020 年度情報処理学会関西支部支部大会講演論文集, C-01(2020).
- [5] 田中達也, 阿多信吾, 村田正幸：ユーザー生成コンテンツの視聴数推移パターン分析と人気推移予測, 信学技報, Vol.116,No.137,IN2016-31,pp.49-54(2016).
- [6] YouTube ヘルプ, チャンネル登録者数を確認する ”<https://support.google.com/youtube/answer/6051134?hl=ja>” (2021.07.09 確認).
- [7] YouTube ヘルプ, YouTube パートナー プログラムの概要と利用資格 ”<https://support.google.com/youtube/answer/72851?hl=ja>” (2021.07.09 確認).
- [8] scikit-learn ”<https://scikit-learn.org/stable/>” (2021.07. 確認).