動画の人気変化に基づくキャッシュ制御の一検討

永田光輝[1]　渋谷彰寿[1]　上山憲昭[2]　山本幹[3]

[1]関西大学大学院理工学研究科
[2]福岡大学工学部
[3]関西大学システム理工学部

動画カテゴリ毎の人気変化

- 動画は各カテゴリ（News, Movie, etc.）で人気変化に差が存在[1]
  - 文献[1]では中国の動画配信サービスであるTencent Video において1日毎の時間粒度で人気変化を観測

- 人気が維持されやすいカテゴリ （MV, TV, Movie）
  - 期間を設定し，その期間中最も要求数の多い動画からキャッシュ
- 人気が維持されにくいカテゴリ（Sport, News）
  - 投稿時刻が最近の動画からキャッシュ

傾向の異なる多様なカテゴリを2種に分類し，
それぞれに異なるポリシーを使用してキャッシュ制御

[1] Y. Zhou, L. Chen, C. Yang and D. M. Chiu, " Video Popularity Dynamics and Its Implication for Replication, " *IEEE TRANSACTION ON MULTIMEDIA*, vol. 7, no. 8, Aug. 2015.

研究目的

既存研究[1]の課題

- 多様なカテゴリに対し，大きく2種の傾向に分類
  - 実際には各カテゴリで傾向は異なる
- 2種ある傾向毎に異なるポリシーを使用
  - 異なるポリシーを使用するためストレージを固定的に分割

考慮すべき点

- 中国の動画配信サービスTencent Videoを対象

研究目的

- YouTubeおけるカテゴリの人気変化の観測
- カテゴリ毎の人気変化に基づくキャッシュ制御の提案

[1] Y. Zhou, L. Chen, C. Yang and D. M. Chiu, " Video Popularity Dynamics and Its Implication for Replication, " *IEEE TRANSACTION ON MULTIMEDIA*, vol. 7, no. 8, Aug. 2015.

YouTubeにおける動画の統計

観測対象

広く普及した動画共有サービスであるYouTube

- 文献[2]を参考にYouTube Data API ver3[3]を用いて測定
  - 動画毎に新規投稿後，1時間毎に視聴回数を取得
  - 取得期間　2017年12月5日~12月25日（20日間）

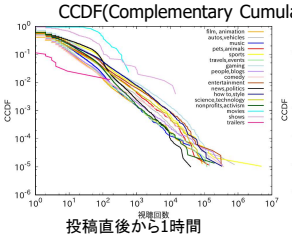
動画をYouTubeが設定している以下の18個のカテゴリで分類

film, animation	gaming	education
autos, vehicles	people, blogs	science, technology
music	comedy	nonprofits, activism
pets, animals	entertainment	movies
sports	news, politics	shows
travels, events	how to, style	trailers

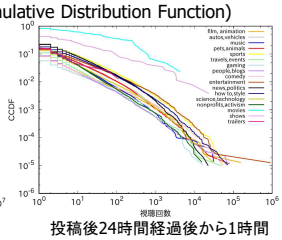
[2] T. Tanaka, S. Ata and M. Murata, " Analysis of Popularity Pattern of User Generated Contents and its Application to Content-aware Networking, " in *Proc. IEEE Globecom Workshop*, Washington, DC, USA, Dec. 2016.
[3] " YouTube Data API ver3, "　https://developer.google.com/youtube/v3/

カテゴリ別視聴回数のCCDF

1時間内に取得した視聴回数の
CCDF(Complementary Cumulative Distribution Function)



投稿直後から1時間



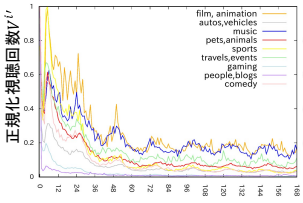
投稿後24時間経過後から1時間

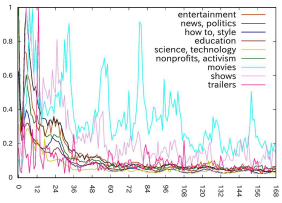
- Movies, Shows
  - 他のカテゴリと比較して，動画全体が視聴される傾向
- 他のカテゴリ
  - 少数の人気な動画に対し，視聴回数が偏る
  - 24時間経過後には多くの動画があまり視聴されない

カテゴリ別人気変化の時間推移

$$V^{i\prime} = \frac{\overline{V_t^i}}{V_{max}^i}$$

$\overline{V_t^i}$：時間[t, t+1]で得られたカテゴリiの平均視聴回数
V_{max}^i：カテゴリiの最大平均視聴回数





- Film animation, Music, Movies, Shows
  他のカテゴリと比較して正規化視聴回数が維持されやすい

カテゴリ毎に人気の変化パターンは異なる

- カテゴリ毎の人気変化は2種に分類できない

提案方式の設計目標

1. 多様なカテゴリの人気変化を考慮
   - 傾向が似たカテゴリでも人気変化には差が存在
   - カテゴリそれぞれの人気変化に対応
2. 単一のポリシーによる制御
   - 人気変化に対し指標を設けることで，全てのカテゴリを同一のポリシーで制御
   - カテゴリ間のキャッシュ領域の共有が容易

各カテゴリが持つ多様な人気変化に基づくキャッシュ制御を提案

人気変化に対する指標

- カテゴリ毎の人気変化に対応する指標としてLife Point(LP)を設定

$$\overline{V_t^i} = \frac{\sum_{j \in C_i} V_t^j}{|C_i|}$$

$$V_{max}^i = \max_{t:0 \to T} (\overline{V_t^i})$$

V_t^j ：時間[t, t+1]で得られた動画jの視聴回数
C_i ：カテゴリiの集合
T：観測終了時刻

$$LP_t^i = \frac{\overline{V_t^i}}{V_{max}^i}$$

カテゴリiの正規化視聴回数の時間推移をカテゴリiの人気変化パターンとして設定

- LPによってカテゴリ毎の多様な人気変化に対応
- 人気が安定するカテゴリは時間経過後もLPを維持しやすい

Life Pointを用いたキャッシュ制御

LPはカテゴリ毎の人気変化パターンを考慮
- カテゴリ内の各動画の人気の差は考慮していない
- カテゴリ内の人気の差を考慮するためV_{t-1}^jを掛ける

V_{t-1}^j: 時間[t−1, t]で得られた動画jの視聴回数
S：ストレージ内のキャッシュの集合
$c(s)$: 動画sのカテゴリ

キャッシュ判断
- カテゴリiの動画j到着時のキャッシュ条件

$$LP_t^i \cdot V_{t-1}^j > \min_{s \in S} (LP_t^{c(s)} \cdot V_{t-1}^s)$$

キャッシュ置換
- ストレージ内最小の$LP_t^{c(s)} \cdot V_{t-1}^s$を持つ動画を削除

LPによる単一ポリシーのキャッシュ制御を実現

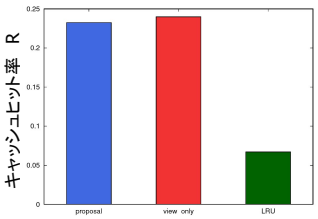
トレースドリブン評価

- シミュレータ　ndnSIM 1.0
- トレースデータ　LP：2017年12月5日～12月12日
  要求発生：2017年12月15日～12月22日
  - この期間に新規投稿された動画を対象
  評価期間：2017年12月16日～12月22日
- 要求発生　トレースデータから取得した視聴回数を到着率としたポアソン分布に従う
- キャッシュサイズ　1000[コンテンツ]
- 評価方式
  - 提案方式(proposal)
  - 1時間前の視聴回数のみでキャッシュ制御(View Only)
  - LRU
- 評価指標
  - 全体のキャッシュヒット率　$R = \frac{\sum_{i=1}^{N} H_i}{\sum_{i=1}^{N} G_i}$
  - カテゴリ毎のキャッシュヒット率　$R_i = \frac{H_i}{G_i}$

H_i: カテゴリiのヒット数
G_i: カテゴリiの要求数
R_i: カテゴリiのヒット率 $R_i = \frac{H_i}{G_i}$
カテゴリ i：1,2,…,N

全体のキャッシュヒット率

キャッシュサイズ 1000[コンテンツ]



H_i: カテゴリiのヒット数
G_i: カテゴリiの要求数
R_i: カテゴリiのヒット率 $R_i = \frac{H_i}{G_i}$
カテゴリ i：1,2,…,N

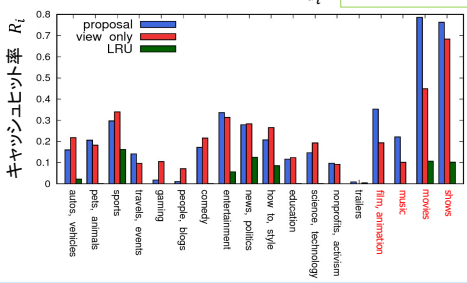
$$R = \frac{\sum_{i=1}^{N} H_i}{\sum_{i=1}^{N} G_i}$$

提案方式はLRUと比較してキャッシュヒット率が向上

カテゴリ別キャッシュヒット率

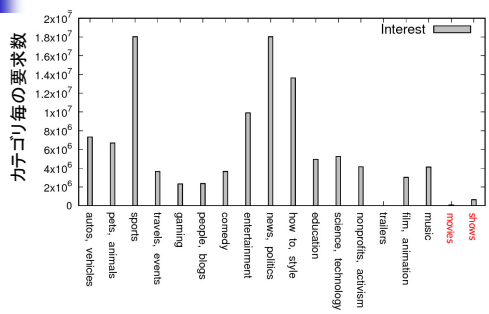
キャッシュサイズ 1000[コンテンツ]　$R_i = \frac{H_i}{G_i}$

H_i: カテゴリiのヒット数
G_i: カテゴリiの要求数



提案方式ではLife Pointを使用することで人気が安定するカテゴリが優位性を持ちキャッシュヒット率が向上

カテゴリ別要求数



提案方式のキャッシュヒット率が高いMoviesやShowは要求数が少ない
➢ 全体のキャッシュヒット率に与える影響が小さい

検討事項

- トレースデータには，MoviesやShowsの動画の一部は含まれていない

MoviesやShowsはYouTubeに認可された公式企業等が投稿
このようなカテゴリは統計情報を開示していない場合が多数存在

- 今後，YouTubeにおけるカテゴリ毎の視聴分布は変化する可能性

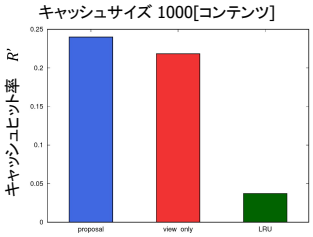
➡ 人気を維持する動画の要求数が増加することで提案方式の性能向上が見込まれる

カテゴリ毎の要求発生数が同一であると仮定した場合の各方式のキャッシュヒット率の評価

均一評価-キャッシュヒット率

キャッシュヒット率$R = \frac{\sum_{i=1}^{N} H_i}{\sum_{i=1}^{N} G_i}$のカテゴリ毎の要求数$G_i$を全て一定$G$に設定

➢ 要求数が同一でもキャッシュヒット率は変わらないと仮定して評価



H_i: カテゴリiのヒット数
G_i: カテゴリiの要求数
R_i: カテゴリiのヒット率 $R_i = \frac{H_i}{G_i}$
カテゴリ i : 1,2,…,N

$$R' = \frac{\sum_{i=1}^{N} R_i}{N}$$

人気が安定するカテゴリの全体の要求数に占める割合が増加することで，提案方式の効果が増加する可能性

まとめ

- YouTubeにおけるカテゴリ毎の人気変化の測定
  - YouTubeにおいてもカテゴリ毎に人気変化の差が存在
- 提案
  - LPを用いたカテゴリ毎の人気変化に対応したキャッシュ制御
- 評価
  - 提案方式のLRUと比較した際のキャッシュヒット率の向上
  - 要求発生数を揃えた際の提案方式の有用性の評価
- 今後の課題
  - 引用論文のキャッシュ制御方式と比較評価
  - シミュレーション期間の延長

参考文献

[1] Y. Zhou, L. Chen, C. Yang and D. M. Chiu, " Video Popularity Dynamics and Its Implication for Replication, " *IEEE TRANSACTION ON MULTIMEDIA,* vol. 7, no. 8, Aug. 2015.
[2] T. Tanaka, S. Ata and M. Murata, " Analysis of Popularity Pattern of User Generated Contents and its Application to Content-aware Networking, " in *Proc. IEEE Globecom Workshop,* Washington, DC, USA, Dec. 2016.
[3] " YouTube Data API ver3, " https://developer.google.com/youtube/v3/

謝辞　本研究の一部は，JSPS科研費17H01740および18K11283により得られたものである．