

Osaka University

Analyzing Popularity Dynamics of YouTube Content and its Application to Content Cache Design

YouTubeコンテンツの人気度推移パタンの分析とコンテンツキャッシュ設計への応用

情報ネットワーク学専攻 村田研究室
北出 雄麻

Osaka University

研究背景

- YouTube に代表される UGC (User Generated Content) の視聴の普及
 - 動画ストリーミング配信は遅延時間に対する要件が厳しい
- 遅延軽減のため、適切なキャッシュ制御手法への期待
 - OTT 事業者 (YouTube など) はネットワーク上の複数箇所に配置したキャッシュサーバを用いてコンテンツを配信
 - ISP が CDN を運用する形態の増加
 - CCN (Content Centric Network) [1] などルータがキャッシュを持つ新たなネットワークアーキテクチャの活発な研究

従来は全ての動画を区別せず、過去のアクセスパターンに基づきキャッシュを制御してきたが、将来の人気度を考慮して制御することが望ましい

[1] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs, and R.L. Braynard, "Networking named content," Proceedings of ACM CoNEXT 2009, pp.1-12, Dec.2009.

Osaka University

研究の目的と手順

- 研究の目的
 - 将来の人気度を考慮したキャッシュ制御手法の実現
- 研究の手順
 - YouTube 動画の日毎の視聴数の収集
 - 時間的・地理的の二つの観点から YouTube の視聴傾向の分析
 - アップロード後の視聴数の推移
 - 大陸, 言語圏, 全世界単位での人気動画の一致度合い
 - 早期の視聴数推移パターンから将来も人気維持されるコンテンツを抽出する分類手法の提案と評価
 - 上記の分類手法を応用し, 将来の人気度を考慮したキャッシュ制御手法の提案と評価

Osaka University

YouTubeの視聴傾向の分析

- 新着動画, 国ごとの人気動画, ランダムに選ばれた動画の日毎の視聴数を YouTube の API を用いて観測
- 視聴数はアップロード後, 日が経つにつれて減少する
- 英語圏の国間の人気動画の一致度合いが高い

アップロード後の視聴数の推移

国間の人気動画の一致度合い

Osaka University

人気の維持される動画の抽出手法の概要

- 将来の視聴数と早期の視聴数には相関がある [9]
- 早期の視聴数の推移パターンに基づき動画を分類し, アップロード直後に視聴数が維持される動画を多く含むクラスタを抽出
 - 各動画 (v 個) について, アップロード後最初の x 日間の視聴数の最大値で各日の視聴数を割り, 得られた v 個のベクトルを用いて k-means 法でクラスタリング
 - アップロード直後に視聴数が維持され, 長期間の人気度が期待される動画を多く含むと予想されるクラスタ $G(k)$ を選出

x = 5 の例

最初の5日間 (x = 5) の最大視聴数100で割る

正規化視聴数

[9] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," Communications of the ACM, vol. 53, pp. 80-88, Aug. 2010.

Osaka University

最適クラスタの決定方法

- アップロード直後に視聴数が維持され, 長期間の人気度が期待される動画を多く含むと予想されるクラスタ $G(k)$ を選出
 - $a_i(k)$: クラスタ数 k のときの, クラスタ i に分類された動画の初期 x 日間の正規化視聴数の重心の値の平均
 - $G(k)$: $a_i(k)$ が最大のクラスタ
- $G(k)$ に分類された動画は人気を持続すると予想

クラスタ数 k = 3, クラスタリングに使用する日数 x = 3 の例

| | | | |
|--|---|--|--|
| <p>クラスタ 1</p> <p>(1, 0.7, 0.3)</p> <p>(1, 0.5, 0.2)</p> <p>(1, 0.6, 0.1)</p> | <p>重心</p> <p>(1, 0.6, 0.2)</p> <p>$a_1(3) = 0.6$</p> | <p>クラスタ 2</p> <p>$a_2(3) = 0.8$</p> | <p>クラスタ 3</p> <p>$a_3(3) = 0.4$</p> |
|--|---|--|--|

G(3): クラスタ 2

Osaka University 7

評価結果 将来の視聴数

- 提案手法により将来の視聴数が多い動画を抜き出せる
 - クラスタ数 k : 15, クラスタリングに使用する日数 x : 30
 - 選択クラスタ ($a_i(k)$ が最大のクラスタ) : クラスタ 1
 - 動画数 : 20,702 (2014/09/06~2015/01/15)

アップロード60日後の視聴数の累積分布関数

視聴数の平均値の推移

Osaka University 8

提案分類手法のキャッシュ制御への応用

- キャッシュ判定
 - 自身がキャッシュしていない動画をキャッシュするかどうかの判断
 - 日の境界時点で、初めて要求が発生してから x 日が経過した動画を対象に提案分類手法を適用
 - 視聴数が維持される動画を多く含むと予想されるクラスタ $G(k)$ の動画と初めて要求が発生してから x 日が経過していない動画は常にキャッシュ
 - それ以外の動画は確率 0.1 でキャッシュ
- 削除動画の選択
 - キャッシュに空きがない際の削除動画の選択
 - LRU (Least Recently Used) を使用

Osaka University 9

評価環境

- コンテンツサーバ、クライアント間にキャッシュ機能を持つルータが1つ存在する場合を想定
- コンテンツ要求の生成
 - YouTube の API を用いて取得したデータ (動画数 : 18,501, 2014/10/06~2015/02/04)
 - 各日の総視聴数の $\frac{1}{100}$ 個の視聴要求を生成
 - 視聴動画はその日の各動画の視聴数に比例する確率で決定
- 比較対象 : 保持していない動画を常にキャッシュ
- キャッシュサイズ : 総動画数の 5% (925 個)
- クラスタリングに使用する日数 x : 10日間
- クラスタ数 k : 15

Osaka University 10

評価結果

- 提案手法によってキャッシュヒット率が向上
 - 人気度上位の動画のヒット率を維持したまま、人気度100番目~1000番目の動画のキャッシュヒット率を向上できている

評価初日からのキャッシュヒット率

動画ごとの相対キャッシュヒット率

Osaka University 11

まとめと今後の課題

- まとめ
 - YouTube の視聴傾向を時間的・地理的の二つの観点から分析
 - 早期の人気度推移パターンから将来も人気維持されるコンテンツを抽出する分類手法の提案
 - 提案分類手法のキャッシュ制御への応用
- 今後の課題
 - キャッシュ制御における最適なパラメータの検討
 - 提案分類手法の他の応用先の検討
 - オンライン広告を掲載する web サイトの絞込