

キャッシュを考慮したコンテンツ推薦システム

1. 研究の背景

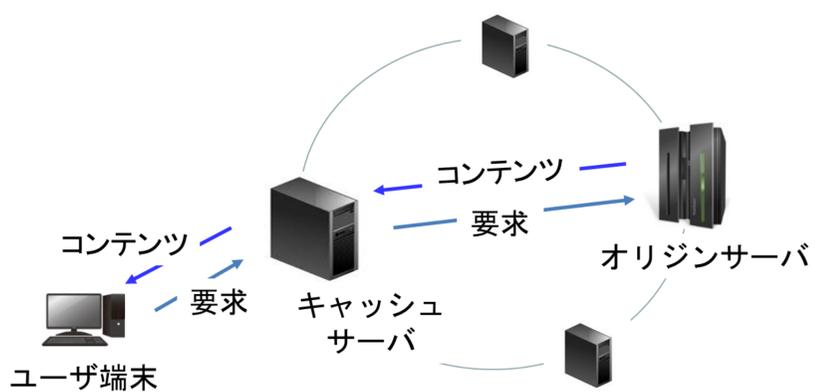
- コンテンツ(ex. 映画, 音楽)の配信サービスの急速な普及
 - コンテンツ数の増加とともに, さらなる需要が予測される
 - 各サービス内のコンテンツの数は膨大
 ⇒コンテンツをユーザ自身が検索し見つけることは困難
- 最適なコンテンツを推薦するシステムが導入
 - ユーザの興味や好みに合ったコンテンツをユーザに提案
- コンテンツプロバイダ(CP: Content Provider)
 - コンテンツを提供する事業者(ex. Netflix, AmazonPrime)
 - 推薦システムがコンテンツ要求の大部分を占有
(Netflixでは約8割を形成)

ユーザの嗜好に沿った推薦

↓

サービスに対する満足度向上

ユーザ数や視聴時間の増加 = CPの収益増加が期待
- ユーザのサービス離脱率
 - QoEを低下させる**配信遅延**などが影響
 ⇒CPはコンテンツを効率的に配信するため、CDNを採用
- コンテンツ配信ネットワーク(CDN: Content Delivery Network)
 - ユーザにコンテンツを効率的に配信するためのネットワーク
 - 地理的に分散配置されたキャッシュサーバで構成
 - ユーザから物理的に近いサーバからコンテンツを配信
 ⇒ネットワークの負荷軽減
⇒コンテンツへのアクセスを高速化



2. 既存の推薦方式

- 協調フィルタリング
 - ユーザの行動履歴(ex. 視聴や購入)を基にした推薦方式
- ① 訪問ユーザAと似た行動履歴をもつ他のユーザBを検索
- ② ユーザBのデータを基に、訪問ユーザAも好む可能性が高いコンテンツをおすすめとして提案

例 Amazon 「この商品を買った人はこんな商品も買っています」

3. 研究の目的

- 従来の推薦では以下の2つが非考慮
 - ネットワーク負荷の軽減
 - 個々のユーザの嗜好
- 研究目的
 - ユーザの満足度とキャッシュの効果向上を目的とした推薦

4. 提案方式の動作

■ キャッシュを活用した協調フィルタリング

例 : 評価点に基づく協調フィルタリングの動作

ユーザの各コンテンツに対する5段階評価を嗜好データとして利用

“-”は未評価

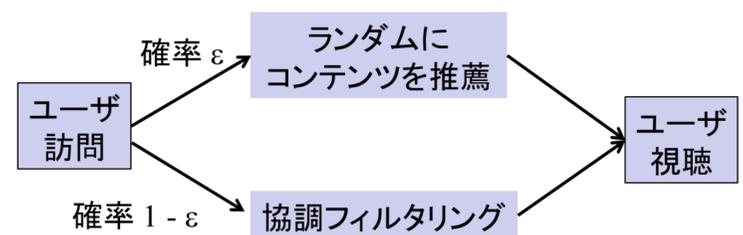
- ① 訪問ユーザXと各ユーザ間で類似度(相関係数)を計算
- ② 相関係数が高い上位k人(例. k=3)の, キャッシュ中のコンテンツ(例. 1,5,6,7)に対する評価度の平均値を計算
- ③ 平均値が最大のコンテンツ(例. コンテンツ5)をユーザXに推薦

キャッシュされているコンテンツ

コンテンツ ユーザ	1	2	3	4	5	6	7	8	9	10	ユーザXとの 相関係数
ユーザX	-	4	2	-	-	-	-	1	1	5	1.000
A	5	3	4	-	-	-	-	1	2	1	0.063
B	-	-	-	1	1	2	4	5	4	0	-0.982
C	2	3	2	1	-	4	3	2	1	4	0.941
D	2	-	-	2	4	5	1	2	5	5	0.500
E	-	5	0	-	5	1	-	1	1	2	0.615
推薦度	2.00				4.50	3.33	2.00				

■ ϵ -greedy法

- 推薦コンテンツのユーザの視聴の有無をデータとして収集
⇒コールドスタート問題(過去のデータが少ない状態)への対応
- 類似度の高いユーザの行動履歴にないコンテンツも推薦
⇒コンテンツの偏りを緩和



■ 推薦コンテンツの多様性

- キャッシュサーバの範囲を拡大
⇒推薦候補のキャッシュコンテンツ数を増加
- キャッシュに限定することによる推薦精度の低下を緩和

