

コンテンツ配信サービスにおける利用者行動に基づくリクエストパターン生成

山崎 重行 奥田 隆史

愛知県立大学大学院 情報科学研究科

Request pattern generation based on user behavior in content delivery network

Shigeyuki YAMAZAKI Takashi OKUDA

Graduate School of Information Science and Technology, Aichi Prefectural University

1 はじめに

動画や音楽などの大容量デジタルコンテンツの配信サービスでは、コンテンツ配信に用途を特定した専用ネットワークであるコンテンツ配信網(以下、CDN: Content Delivery Network)が利用されている。CDNで利用されている技術の特色は、ネットワーク上の様々な場所にコンテンツの配布ポイントを設置し、そこにコンテンツを複製配置することである。このような技術により多数の利用者に配信が可能になる。ISP(Internet Service Provider)などはCDNを円滑に運営し、利用者に良質なサービスを提供する必要がある。そこでネットワークの性能評価が重要になる。

性能評価を実施するためには、利用者からCDNへのリクエストを把握する必要がある。これまで利用者からコンテンツに対するリクエストの分布(以下、リクエスト分布)はZipf分布に従うとし、数理モデルやシミュレーションによって性能評価が行われている。しかし、動画や音楽などのコンテンツのリクエスト分布はZipf分布では表現できないと指摘されている[1]。適切な性能評価を実施するためには、実際のリクエスト分布に即した評価が必要である[2]。

そこで本研究では、リクエストをネットワークと利用者の相互作用と捉えて把握するアプローチにより、CDNの性能評価を実施する。この種の相互作用を本研究ではマイクロマクロリンクと呼ぶ。マイクロマクロアプローチにより例えば、新しいコンテンツを提供した場合、どの程度リクエストが発生し、また、すでに提供されているコンテンツにどの程度影響が表れるかを考慮した性能評価が可能になる。

マイクロマクロリンクアプローチによる性能評価を実現するため、複数のエージェントによりマイクロマクロリンクを表現できる、Agent Base Modeling(以下、ABM)を用いる。ABMによりCDNにおける利用者やコンテンツ、配布ポイントをエージェントと捉え、CDNの性能評価を行う評価シミュレータを開発する。また、利用者の行動をモデル化し、評価シミュレータによりそのモデルによるリクエストパターンの生成を行う。

以下、2.節では、本研究で想定するCDNと利用者について述べ、3.節でそれらをエージェントとして捉えモデリング手法を示す。4.節では利用者行動モデルについて説明し、5.節ではそのモデルによるリクエストパターンの作成を行う。最後に6.節でまとめる。

2 CDNと利用者

本研究で想定するCDNネットワーク構成と利用者を、図1に示す。以下ネットワーク構成と利用者のリクエスト行動について述べる。

2.1 ネットワーク構成

CDNのネットワーク構成は、3階層のネットワークとする。図1の右側より、最上位層(higher-level layer)には、オリジナルのコンテンツを持つOs(Origin server)がある。中間層(Intermediate layer)には、コンテンツをOsより他の配布ポイントに配信するDs(Distribution System)がある。最下位層(Lower-level layer)には、コンテンツを利用者にOsの代理となって配信を行うSg(Surrogate)がある。Sgは、ISPごとに1台設置されていると想定する。

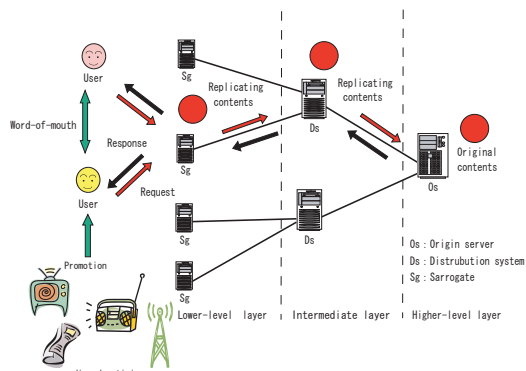


図1 CDNネットワーク構成と利用者

利用者からのリクエストをどの配布ポイントに割り振るか決定するリクエストルーティングは、利用者からのリクエストを、あらかじめ決められているSgに誘導すると仮定する。これは、利用者が所属しているISPのネットワーク内のSgに割り振られることを想定している。

コンテンツ配信方法は、配信アルゴリズムに従い、複製を配布ポイントに配信する。Ds, SgがOsより読み出したコンテンツを保存する時、自身のコンテンツ保存容量を超過する場合は、置き換えアルゴリズムに従いコンテンツを削除する。

2.2 利用者のリクエスト行動

利用者がCDNにリクエストを行う行動は、CDNに存在しているコンテンツを知覚し、その中より興味のあるコンテンツを選択してリクエストを決定するものとする。

コンテンツの選択とリクエストの決定は、

- 利用者の嗜好。
- ISPやコンテンツ配給会社などの企業が行うプロモーション活動。
- 利用者同士の口コミ。
- コンテンツのコストと支払う支払意思額(WTP: Willingness To Pay)。
- リクエストを出してからコンテンツ入手までの待ち時間。

の要因で影響を受けると仮定する。また、利用者は多様であるとして、上記の影響は利用者によって異なるとする。

3 マイクロマクロリンクアプローチによるモデリング

本研究では、利用者及びCDNの配布ポイントとコンテンツをマルチエージェントシステムとして捉え、ABMにより表現する。マルチエージェントシステムSを構成するエージェントと環境は、

$$S = \{Environment, Agent\}$$
$$Agent = \{Osa, Dsa, Ssa, Ua, Ca\}$$

と表現する。Environmentは、コンテンツに対するプロモーションとする。Agentは5種類とし、OsaはOrigin server, DsaはDistribution system, SgaはSurrogateを表す。また、Uaは利用者を、Caはコンテンツを表す。

CDNによるコンテンツ配信サービスを表現するために、エージェント間でメッセージのやり取りを行う。その結果、エージェントや環境に影響を与える。これらを相互作用とし、図2に破

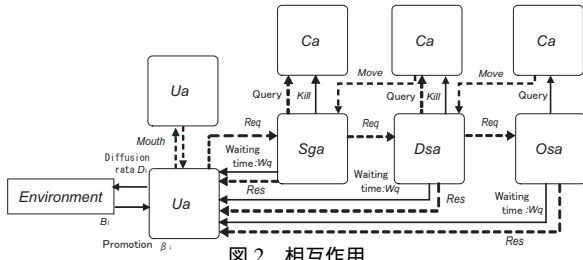


図2 相互作用

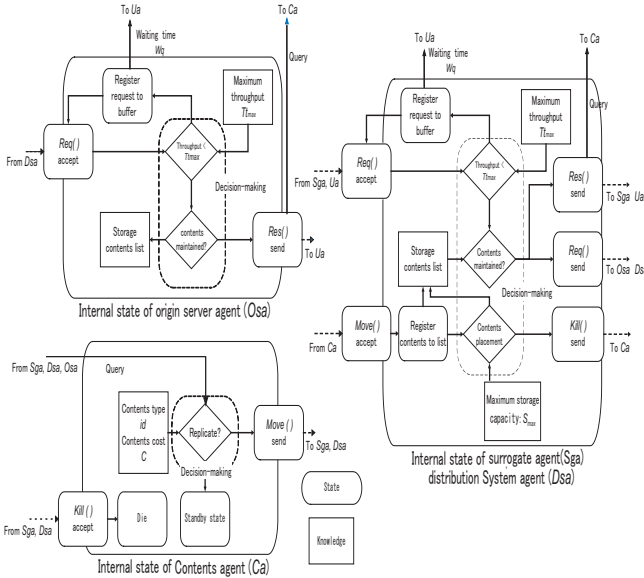


図3 エージェントの内部状態と知識

線をメッセージ、実線を影響として示す。また、 Ua 以外のエージェントの内部状態と知識を図3に示す。 Ua の内部状態と知識については、次節で述べる。

このようなマルチエージェントシステム S により、シミュレーションを実施することで、CDN の性能評価を行う評価シミュレータを開発する [3]。評価シミュレータ開発には、プラットフォームに KK-MAS を用いる。

4 利用者行動のモデリング

2.2 節で述べた利用者のリクエスト行動を表すため、利用者エージェント Ua の意思決定と内部状態を図4のようにする。コンテンツの選択やリクエストの決定に与える影響は相互作用により表す。また、コンテンツの選択やリクエストの決定は、利用者行動モデルとして扱う。利用者行動モデルは、図4の破線で囲まれた範囲に示す5つの過程でモデル化できるものとする。それぞれの過程は、

- 情報取得 (Information gathering)
- コンテンツ選択 (Selecting content)
- リクエスト決定 (Request decision)
- コンテンツ評価 (Evaluating content)
- 口コミ行動 (Word-of-mouth behavior)

である。以下それぞれの過程について説明する。

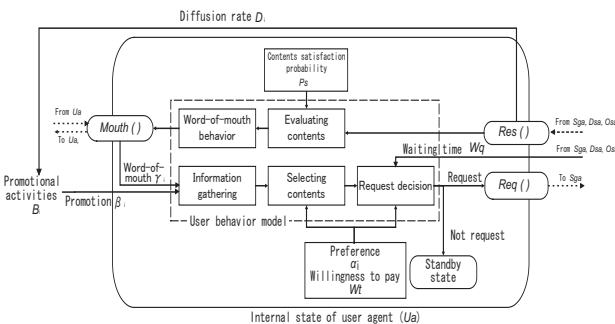


図4 利用者行動モデル

4.1 情報取得

情報取得では、利用者 $j(j = 1, 2, \dots, J)$ がコンテンツ $i(i = 1, 2, \dots, I)$ を選択するための情報を取得する。このような情報をコンテンツ情報 A_i^j とし、4.1.1 節で説明する。また、取得方法については、4.1.2 節で述べる。

4.1.1 コンテンツ情報

時刻 t において、利用者 j がコンテンツ i に対して持つコンテンツ情報 $A_i^j(t)$ は、

$$A_i^j(t) = \{\alpha_i^j(t), \beta_i^j(t), \gamma_i^j(t)\} \quad (1)$$

とする。 $\alpha_i^j(t)$ は、 j が i に対する好みの程度を表し、この値が大きい程 j が i に対してリクエストする可能性が高くなる。 $\beta_i^j(t)$ は、 i に対してのプロモーションに接触した回数を表す。プロモーションに接触する確率をプロモーション活動量 $B_i(t)$ とする。

口コミ回数 $\gamma_i^j(t)$ は、 j が i に対して他の利用者より受けた口コミの回数を表す。口コミとは、コンテンツの内容について他の利用者にリクエストを薦める、または薦めない情報である。

なお本研究では、 $\beta_i^j(t)$ と $\gamma_i^j(t)$ の回数による $\alpha_i^j(t)$ への影響は、簡単のために同等と仮定する。また $A_i^j(t)$ は、取得してから一定期間経過後、利用者の記憶を表現するため、利用者から削除されるとする。この期間を記憶時間 T_p とする。

4.1.2 コンテンツ情報の取得方法

コンテンツ情報 $A_i^j(t)$ の取得は、口コミとプロモーションからと想定する。口コミによる $A_i^j(t)$ の取得は、他の利用者より $A_i^j(t)$ について伝達される時である。

プロモーションによる $A_i^j(t)$ を取得する確率をプロモーション活動量 $B_i(t)$ と定義する。 $B_i(t)$ は、企業の i に対するプロモーション規模に影響されると考え、企業の宣伝広告費やプロモーション期間などにより決定されると想定する。また、インターネットにおける不特定多数の書き込みなどもこのようなプロモーションに該当すると考える。そのため、コンテンツ普及率 $D_i(t)$ が高くなると、多数の利用者より書き込みなどが多くなると仮定し、その結果 $B_i(t)$ が増加するとする。

4.2 コンテンツ選択

コンテンツ選択過程では、利用者 j はコンテンツ情報 $A_i^j(t)$ に基づき確率的にコンテンツ i を選択するものと仮定する。この確率を文献 [4] のブランド選択モデルを用いて算出する。

時刻 t において j が i に対しリクエストする選択確率 $P_i^j[A_i^j(t)]$ は、

$$P_i^j[A_i^j(t)] = \frac{\exp\{V_s[A_i^j(t)]\}}{\sum_{m \in I} \exp\{V_s[A_m^j(t)]\}} \quad (2)$$

$$V_s[A_i^j(t)] = \alpha_i^j \cdot (1 - \exp[-a^j \cdot (\beta_i^j(t) + \gamma_i^j(t))]) \quad (3)$$

である。パラメータ $a^j(a > 0)$ は定数であり、 j のプロモーションと口コミに対する影響度を表す。効用の最大値は $\alpha_i^j(t)$ に依存するとする。

4.3 リクエスト決定

リクエスト決定過程では、4.2 節より選択したコンテンツのリクエストを行うか決定する。この決定については、文献 [4] の購買生起モデルを用いる。

時刻 t において利用者 j が、選択したコンテンツ i に対しリクエストするリクエスト確率 $P_R^j[A_i^j(t)|i]$ は、

$$P_R^j[A_i^j(t)|i] = \begin{cases} \frac{1}{1 + \exp\{\Phi \cdot G[A_i^j(t)]\}} & G[A_i^j(t)] \geq 0 \\ 0 & G[A_i^j(t)] < 0 \end{cases}$$

となる。 $G[A_i^j(t)]$ は j が i をリクエストした場合の利得を表す。 $G[A_i^j(t)]$ が 0 以上ならリクエストの意思決定を行い、0 より小さければリクエストを行わない。 Φ は利得のパラメータである。

利得 $G[A_i^j(t)]$ は、

$$G[A_i^j(t)] = W[A_i^j(t)] - C_i \quad (4)$$

$$W[A_i^j(t)] = \omega^j \cdot V_s[A_i^j(t)] \cdot Wq^j(t) \cdot W_i^j \quad (5)$$

とする。\$C_i\$ はコンテンツのコストを表す、コンテンツ効用 \$V_s[A_i^j(t)]\$ は、式 (3) を用いる。\$W_i^j\$ は \$j\$ の WTP である。\$\omega^j\$ (\$\omega^j > 0\$) は、\$V_s[A_i^j(t)]\$ が \$W[A_i^j(t)]\$ に与える影響度である。\$Wq^j(t)\$ は、

$$Wq^j(t) = \exp[-\sigma^j \cdot T_q^j(t)] \quad (6)$$

とし、\$j\$ の平均待ち時間 \$T_q^j(t)\$ が \$W[A_i^j(t)]\$ に与える影響度である。式 (4) により \$P_{Ri}^j[A_i^j(t)|i]\$ を算出し、その確率に従いリクエストを決定する。

4.4 コンテンツ評価

コンテンツ評価では、リクエスト過程で取得したコンテンツ \$i\$ の内容に対し満足するかしないかの評価を行う。この満足する確率を満足確率 \$P_s\$ とする。\$i\$ に対して満足した場合は、他の利用者にリクエストを勧める情報を持つ。これを正の口コミと呼ぶ。一方、満足しない場合は、リクエストを薦めない情報を持つ。これを負の口コミと呼ぶ。

4.5 口コミ行動

口コミ行動過程では、前項の正の口コミか負の口コミを他の利用者へ伝える。

口コミ行動を表現するために、利用者 \$j\$ が口コミを伝達する範囲を、図 5 のような論理的な利用者同士の対人関係を表した空間上において設定する。

なお、空間はループするものとし、空間の端に存在する利用者は、反対側の端の利用者と繋がっているものとする。利用者同士の親密度の大きさを表すため、対人距離 \$d\$ (Degrees of separation) を定義する。例えば、\$d = 1\$ なら、\$j\$ にとって最も親密な関係にある他の利用者であり、家族や親友などと考える。\$d\$ が大きくなるに従い、\$j\$ と他の利用者の親密度は小さくなる。本研究では、\$d = 1\$ とし、最大で口コミを行う人数は 4 人とする。

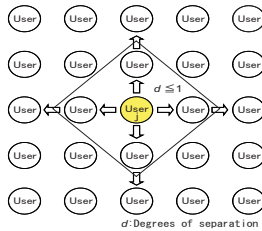


図 5 口コミの範囲

5 リクエストパターン生成

本節では、前節で述べた利用者行動モデルを利用者エージェント \$U_a\$ に導入し、\$U_a\$ の知識やパラメータを変更した場合のリクエストパターンを生成するために、評価シミュレータの設定例を用いてシミュレーションを実施する。

シミュレーションは、60 日間行うものとし、各エージェントの知識の設定は、表 1 とする。また、利用者行動モデルの設定例を表 2 とする。

プロモーション活動量 \$B_i(t)\$ の値は、

$$B_i(t) = \frac{\text{コンテンツ認知率}}{\text{シミュレーション時間}} + D_i(t)^2 \quad (7)$$

により決定する。コンテンツ認知率とは、シミュレーション終了時に利用者全体に対してコンテンツ \$i\$ の情報を取得した経験がある利用者数の割合である。設定例では、\$B_i(t)\$ は、コンテンツ認知率を区間 \$[0,1]\$ の一様分布で決定するものとし、コンテンツ普及率 \$D_i(t)\$ の 2 乗の影響を受けて増加するとする。また、利用のパラメータ \$\Phi = 0.1\$ とする。

それ以外の評価シミュレータでの仮定条件は、以下のようにする。

- コンテンツ \$i\$ の転送時間やデータサイズについては考慮しない。
- 24 時間ごとに新しいコンテンツを \$O_s\$ へ追加する。
- 利用者 \$j\$ は、一度リクエストして取得した \$i\$ を再度リクエストしない。

表 1 エージェントの設定例

知識		入力値	
\$Sga\$	最大処理量	\$Th_{maxSg}\$	50 メッセージ
	最大保存量	\$S_{maxSg}\$	20 種類
\$Dsa\$	最大処理量	\$Th_{maxDs}\$	100 メッセージ
	最大保存量	\$S_{maxDs}\$	50 種類
\$Osa\$	最大処理量	\$Th_{maxOs}\$	200 メッセージ
\$Ca\$	コンテンツの種類	\$ID\$	100 種類
	コンテンツコスト	\$C_i\$	10
\$Ua\$	コンテンツに対する嗜好	\$\alpha_i^j\$	区間 \$[0,10]\$ に従う一様分布
	支払い意思額	\$Wt^j\$	平均 10、分散 1 に従う正規分布
	コンテンツ満足確率	\$P_s^j\$	1.0

表 2 利用者行動モデルの設定例

利用者行動モデル	パラメータ	記憶期間 \$T_p\$	1 時間
	\$\alpha^j\$	区間 \$[0,1]\$ に従う一様分布	
	\$\sigma^j\$	区間 \$[0,1]\$ に従う一様分布	
	\$\omega^j\$	0.2	

シミュレーション結果として、乱数を変更し、3 回シミュレーションを実施した値の平均値を用いる。リクエスト状況を図 6 に、そのリクエスト分布を図 7 に示す。

図 6 は、シミュレーション開始時から 30 日目にプロモーション活動量 \$B_i = (0.0023, 0.0009)\$ のコンテンツを \$O_s\$ に配置した場合の、1 日あたりのリクエスト数の推移の一例を示す。これらの \$B_i\$ の値は、コンテンツ認知率を 50%、20% として算出した。図 6 より、1 日あたりのリクエスト数が増加して減少するパターンと徐々に増加していくパターンが在ることが分かる。

図 7 は、コンテンツのリクエスト数の多いものから並び替えたものを横軸とし、そのリクエスト累積確率を縦軸として示す。利用者行動モデルによるリクエスト分布と既存の性能評価手法でリクエスト分布として用いられている Zipf 分布の比較のため、\$\alpha = 0.415\$ の Zipf 分布を図 7 に示す。この \$\alpha\$ は、統計解析ソフト R[5] の非線形回帰分析 (最小 2 乗法) を用いて求めた。

図 7 より、リクエスト数が中程度のコンテンツに対しては Zipf 分布よりもリクエスト数が多いことが分かる。

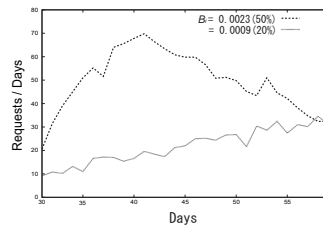


図 6 リクエスト状況

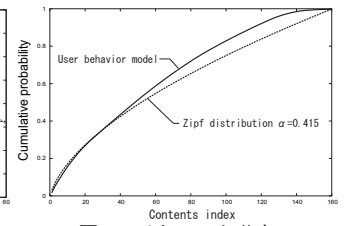


図 7 リクエスト分布

6 おわりに

本研究では、CDN と利用者、コンテンツをエージェントと捉え ABM により表現し、性能評価を実施するためのシミュレータを開発した。また、利用者の行動のモデリングを行い、ミクロマクロリンクによるリクエストパターンを生成することが可能となった。このようなリクエストを用いて、CDN におけるコンテンツ配置問題の評価に利用することができる [6]。

今後の課題としては、実際の CDN における利用者からのリクエストパターンとの比較や、利用者行動モデルのパラメータの検討が挙げられる。

参考文献

- [1] C. Griwodz, M. Bar, and L. C. Wolf, "Long-term movie popularity models in video-on-demand systems", In *Proceedings of ACM Multimedia*, 1997.11.
- [2] 山崎重行, 奥田隆史, 井手口哲夫, 田学軍, "CDN における利用者行動に基づくリクエスト分布の評価", 情報学ワークショップ 2005(WiNF2005), pp.125-130, 愛知県立大学, 2005.9.
- [3] 山崎重行, 奥田隆史, 井手口哲夫, "マルチエージェントによる CDN システムの性能評価手法", 信学技報 (IN2005-27), pp.77-82, 愛知県立大学, 2005.6.
- [4] 岡太彬訓, 木島正明, 守口剛, マーケティングの数理モデル, 朝倉書店, 2002.
- [5] The R Project for Statistical Computing, <http://www.r-project.org/>.
- [6] 山崎重行, 奥田隆史, 井手口哲夫, "マルチエージェントモデルを用いたコンテンツ配信サービス網におけるコンテンツ配置アルゴリズムの評価", 合同エージェントワークショップ & シンポジウム 2005(JAWS2005), pp.305-312, 箱根ホテル小涌園, 2005.10.