

[ポスター講演] キャッシュ制御のための Web オブジェクト共起度分析

桜井 沈輝[†] 上山 憲昭[†]

[†] 福岡大学工学部電子情報工学科 〒814-0180 福岡市城南区七隈 8-19-1
E-mail: ††1151225@cis.fukuoka-u.ac.jp, ††kamiyama@fukuoka-u.ac.jp

あらまし Web ページの表示待ち時間を低減する技術として、Web ページを構成する複数のオブジェクトを同時並列に取得する HTTP/2 が標準化され広く普及している。しかし HTTP/2 の並列配信は同一の配信サーバから取得するオブジェクトに対してのみ可能であり、Web 表示待ち時間の低減効果を高めるには、少数の配信サーバから多数のオブジェクトを取得する必要がある。そこであるオブジェクトの組に対し、その組が出現する Web ページの数を共起度と定義すると、共起度の高いオブジェクト組が優先的にキャッシュに残るよう、キャッシュ置換を行うことが望ましい。しかし共起度に基づくキャッシュ置換制御の効果は、実際の Web ページにおいてどの程度、オブジェクト間の共起現象が生じるかに依存する。そこで本稿では、共起度に基づくキャッシュ置換制御の可能性を明らかにするため、高人気の 8,000 の Web ページを構成する約 68 万個のオブジェクトを対象に、オブジェクトの 2 個組と 3 個組の共起度を計算し、Web ページの共起現象の程度を分析する。その結果、共起度の分布は冪乗則に従い、0.1% 程度の 2 個組や 0.5% 程度の 3 個組は 100 以上の、0.005% 程度の 2 個組や 0.001% 程度の 3 個組は 500 以上の共起度を有することを明らかにする。そのためこれら少数の共起度の高いオブジェクト組を優先的にキャッシュに残すキャッシュ置換制御の有効性が期待される。

キーワード HTTP/2, Web オブジェクト, 共起

[Poster Presentation] Measurement Analysis of Co-occurrence Degree of Web Objects for Web Cache Control

Kouki SAKURAI[†] and Noriaki KAMIYAMA[†]

[†] Department of Electronic and Information Technology, Fukuoka University 8-19-1, Nanakuma, Jounan, Fukuoka 814-0180
E-mail: ††1151225@cis.fukuoka-u.ac.jp, ††kamiyama@fukuoka-u.ac.jp

Abstract As a technique to reduce the web response time, HTTP/2 which enables user terminals to simultaneously download multiple web objects consisting one webpage has been standardized and widely used. However, parallel download of HTTP/2 is possible only for web objects which are provided from the same object or cache server, so a large number of objects need to be downloaded from a small number of servers to improve the effect of HTTP/2. Therefore, it is desirable to replace cache servers so that object sets with high co-occurrence degree are remained in the cache, where we define the co-occurrence degree of object set as the number of webpages in which the object set is included. However, the effect of cache-replacement method based on the object co-occurrence degree depends on how the co-occurrence phenomenon of objects appears in the actual webpages. Therefore, to clarify the possibility of cache-replacement method based on the co-occurrence degree, we investigate the degree of co-occurrence phenomenon of objects by calculating the co-occurrence degree of two-objects and three-objects pairs of about 0.68 million objects constructing popular 8,000 webpages in this paper. We confirm that the distribution of co-occurrence degree of web objects obeys the power law. We also clarify that about 0.1% two-object pairs and 0.5% three-object pairs have more than 100 co-occurrence degree, and about 0.005% two-object pairs and 0.001% three-object pairs have more than 500 co-occurrence degree. Therefore, we can expect the effectiveness of replacing cached objects so that a small number of object sets with extremely high co-occurrence degree are remained with high priority.

Key words HTTP/2, web object, co-occurrence

1. はじめに

Web 閲覧サービスは、インターネットにおける主要サービスの 1 つであり、多くの人が日常的に利用している。しかし 2/3 のユーザは毎週のように Web サイト閲覧時の低速性を経験しており [2], 17% のユーザは待ち時間が 5 秒を超えた場合には閲覧を諦めるといった報告もなされている [3]。また 400 ミリ秒の遅延により Google サーチエンジンでの検索回数が 0.74% 低減することや [11], Web 応答時間が 0.1 秒だけ削減することにより Amazon の利益が 1% 増加することが報告されている [12]。また高速に表示される Web サイトはユーザが購買を完了する回数が 15% も多く、また 1 ページだけ閲覧した後にサイトから離脱する回数が 9% も少ないことが報告されている [8]。そのためインターネットの主流サービスである Web トラフィックを適切に制御することが、ユーザの体感品質を向上させ、ネットワーク資源の消費量を抑えるためには重要である。

従来の Web サイトは静的なテキストや画像といったオブジェクトがサーバに用意され、Web ブラウザは HTTP を用いてこれら静的オブジェクトを単にダウンロードして表示していた。しかし近年、クライアント端末からの要求受信時に、サーバ側で実行するか、JavaScript で書かれた Ajax や DOM (document object model) によるプログラムを HTML に埋め込みユーザ端末側で実行することで生成される動的オブジェクトの割合が増加している [2]。また、広告を専用のサーバから取得するなど、各オブジェクトの配信元が多様化している。このように一つの Web サイトを構成するオブジェクトは複雑性を増している。

Web オブジェクトの取得には TCP セッション上で HTTP (hypertext transfer protocol) が用いられる。従来、広く用いられていた HTTP/1.1 では、オブジェクトの取得に要する遅延時間を低減するため、同一配信ホストから取得する複数のオブジェクトを一つの TCP セッション上で取得する HTTP persistent connection と、さらに同一ホストから複数のオブジェクトを並列に取得する HTTP pipelining が実装されている。しかし同一の TCP コネクション上で転送されるパケットが属するオブジェクトをユーザ端末が識別できないため、配信サーバは HTTP request を受信した順番でオブジェクトを返信する必要があり、送信準備ができたオブジェクトの返信開始が、他のオブジェクトの返信完了まで待たされる Head of Line (HOL) 問題が生じる。図 1(a) に HTTP/1.1 を用いて a, b, c の 3 個オブジェクトを同一の配信ホストからユーザ端末がダウンロードする場合の配信フローを例示する。図では 3 つのオブジェクトに対する配信要求が a, b, c の順で配信ホストに到着し、オブジェクト b の生成に時間を要した場合を示している。オブジェクト c の配信準備は完了しているが、配信サーバは a, b, c の順オブジェクトを送信する必要があるため、オブジェクト b の送信が完了した後で、オブジェクト c の配信が可能となる。

このような HOL 問題を解決するため Google は、パケットに「SPDY stream」と呼ばれる所属 HTTP セッションの識別子を付加することで、ユーザ端末が各パケットの所属オブジェ

クトを識別可能とし、配信サーバが任意の順番でオブジェクトを配信可能とする SPDY を開発した [6] [12] [13]。図 1(b) に、3 つのオブジェクトを SPDY で配信した場合の例を示す。配信サーバは任意の順番でオブジェクトの配信が可能であり、先に配信準備が完了したオブジェクト c をオブジェクト b に先立ち送信を開始できる。SPDY は HTTP の機能として標準化され、現在、SPDY が組み込まれた HTTP が HTTP/2 として普及しつつある [7]。新バージョンである HTTP/2 にはいくつかの機能が追加され、その中でも特に主要な機能である SPDY による多重化やサーバプッシュ機能により HTTP/1.1 までの問題点であった HOL 問題が回避され、Web ページの読み込み速度への大きな改善が期待されている。

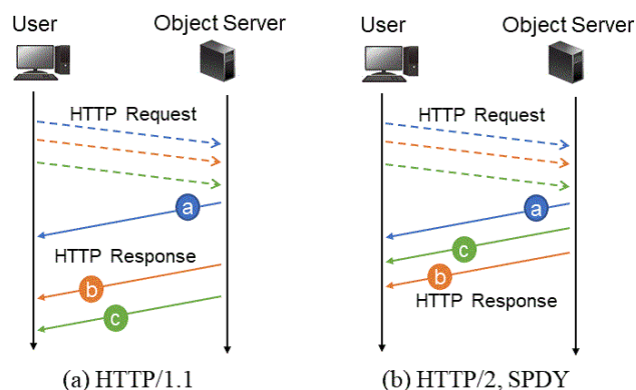


図 1 Example of object delivery sequence in HTTP/1.1 and HTTP/2 (SPDY)

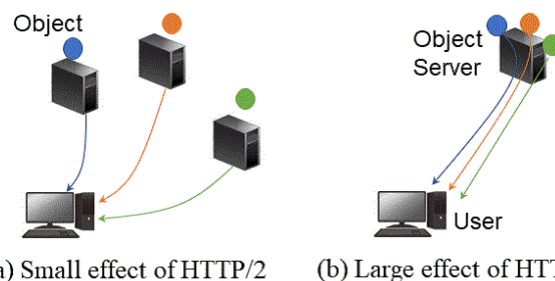


図 2 Comparison of two cases of delivering three objects from three servers or one server

しかし HTTP/2 のオブジェクト並列配信機能は、同一の配信サーバから取得するオブジェクトに対してのみ有効である。すなわち異なる配信サーバから取得するオブジェクトを同一 TCP セッション上で取得することができず、HTTP/2 の並列配信による遅延時間削減効果も期待できない。例えば図 2 に 3 個のオブジェクトを 1 個ごとに別のサーバから取得する場合と、同じサーバから取得する場合の例を示すが、HTTP/2 の効果は同一の配信サーバから取得するオブジェクト数が増えるほど大きくなる。そのため HTTP/2 の効果は期待したほど得られないという報告もなされている [4]。Web 表示待ち時間の低減効果を高めるには、少数の配信サーバから多数のオブジェクトを取得する必要がある。

そこであるオブジェクトの組に対し、その組が出現する Web ページの数を共起度と定義すると、共起度の高いオブジェクト組が優先的にキャッシュに残るよう、キャッシュ置換を行うことが望ましい。しかし共起度に基づくキャッシュ置換制御の効果は、実際の Web ページにおいてどの程度、オブジェクト間の共起現象が生じるかに依存する。そこで本稿では、共起度に基づくキャッシュ置換制御の可能性を明らかにするため、実際の Web ページを構成するオブジェクトを対象に共起度を計算し、Web ページの平均共起度などの共起度に関する各種特性を分析する。以下 2 節では共起度に基づくキャッシュ制御方式の概要を述べる。そして 3 節では、実際の Web ページにおける共起度の測定分析結果について述べ、最後に 4 節で全体をまとめる。

2. HTTP/2 の効果向上のための共起度に基づくキャッシュ制御

1. 節で述べたように HTTP/2 の並列配信の効果は同一の配信サーバから取得するオブジェクト集合に対してのみ有効である。そのため多数の配信サーバからオブジェクトを、各々からは少数だけ取得する場合には HTTP/2 の効果は低減する。ところで Web 応答時間を低減する技術として CDN (Content Delivery Network) が広く普及している [8] [9] [10]。CDN はネットワーク上の多数の場所に設置されたキャッシュサーバにオブジェクトのコピーをキャッシュし、ユーザの近くに存在するキャッシュサーバからオブジェクトを配信することで Web 応答時間を低減する。HTTP/2 を利用するには配信サーバとユーザ端末の両方が対応する必要があるが、オブジェクトのオリジナルを提供するオリジンサーバが HTTP/2 に未対応であっても、CDN 事業者がキャッシュサーバを HTTP/2 に対応させることでキャッシュサーバとユーザ端末間で HTTP/2 を利用することができる。既に Akamai、Amazon CloudFront、Microsoft Azure、Google Cloud Platform などの大手 CDN 事業者が HTTP/2 に対応している。

キャッシュサーバは配信要求に対しキャッシュに要求オブジェクトが存在しない場合、オリジンサーバからオブジェクトを取得してキャッシュしたのちユーザに配信する。キャッシュの空き容量が不足する場合はキャッシュ済みオブジェクトの一部を削除するが、その選択方法 (キャッシュ置換法) がキャッシュの Web 応答時間の低減効果に影響する。本稿では任意の 2 つのオブジェクトが同一の Web ページ内で出現することを「共起」と呼び、同一の Web ページ内で共起する他のオブジェクトの数を「共起度」と呼ぶ。例えばあるオブジェクト組を含む 20 の Web ページが存在する場合、これらオブジェクト組の共起度は 20 となる。また図 3 に青、オレンジ、緑の 3 つのオブジェクトから構成される 3 つの Web ページが存在する状況における、オブジェクトの 3 つの組の共起度を例示する。例えば青とオレンジのオブジェクト組は、2 つの Web ページに出現しているため共起度は 2 となる。

従来のキャッシュ制御方式は各オブジェクトと独立に扱う。しかし HTTP/2 の並列配信による Web 応答時間の低減効果

は、共起度の高いオブジェクトほど大きくなる。そのため共起度の高いオブジェクトの集合が優先して残るよう、キャッシュ置換を行うことが望ましい。例えば共起度の高い 10 個のオブジェクトの組が存在するとき、これら 10 個のオブジェクトの組を一塊として、キャッシュへの挿入と削除を行うことが考えられる。その結果、1 つの Web ページを閲覧したときに同一のキャッシュサーバから配信されるオブジェクト数が増加し、HTTP/2 の効果の向上が期待される。

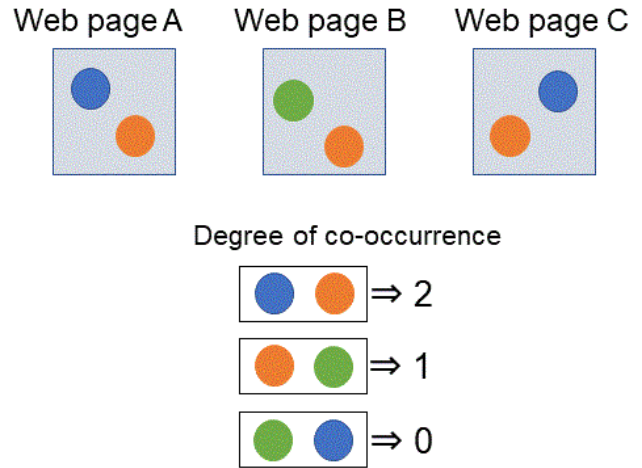


図 3 Degree of co-occurrence of three two-objects pairs among three web pages

3. 共起度の測定分析

2. 節で述べた共起度に基づくキャッシュ置換制御の効果は、現実の Web ページにおいてどの程度、共起現象が見られるかに大きく依存する。そこで本稿では、実際の多数の Web ページに含まれるオブジェクトの共起度を測定分析することで、共起度に基づくキャッシュ制御の可能性を明らかにする。

3.1 オブジェクトデータの測定法

本節では、Web ページに含まれるオブジェクトデータの測定手順を述べる。アクセス数の多い高人気の Web ページに含まれるオブジェクトを共起度分析の対象とすることが望ましい。そこで adult, arts, business, computers, games, health, home, kids&teens, news, recreation, reference, regional, science, shopping, society, sports の 16 のカテゴリごとに、Web ページのアクセス数のランキングを公開している Alexa の Web ページから [1], 各カテゴリに対して上位 500 の Web ページの URL を測定対象としてリスト化した。次に生成した評価 URL リストの各 URL に対して、測定用クライアント PC から GET の HTTP リクエストを送信した際に発生する通信特性を、HAR (HTTP Archive) ファイルとして取得した [5]。HAR ファイルは、クライアント PC とサーバ間で転送される HTTP データのヘッダ情報から、クライアント PC において、各オブジェクトの URL、サイズ、取得に要した遅延時間等の各種通信特性を算出し、JSON (JavaScript Object Notation) 形式で出力したものである。JSON ファイルを解析することで、各 Web ペー

ジを構成する各オブジェクトの URL 名を取得し、URL 名で全てのオブジェクトを識別した。

3.2 オブジェクトの重複度

16 の各カテゴリから 500 の合計で 8,000 の Web ページを閲覧した結果、取得したオブジェクトの総数は 679,380 個であった。各オブジェクトに対し、8,000 の Web ページの中で出現する Web ページの数を重複度と定義する。図 4 に、679,380 個の各オブジェクトの重複度の累積補分布 (CCD: complementary cumulative distribution) を示す。重複度 1 のオブジェクトが全体の 9 割程度を占めるが、重複度の分布の裾野は広く、冪乗則が観測される。0.05% 程度のオブジェクトは 100 以上の Web ページに、0.001% 程度のオブジェクトは 1,000 以上の Web ページに出現している。

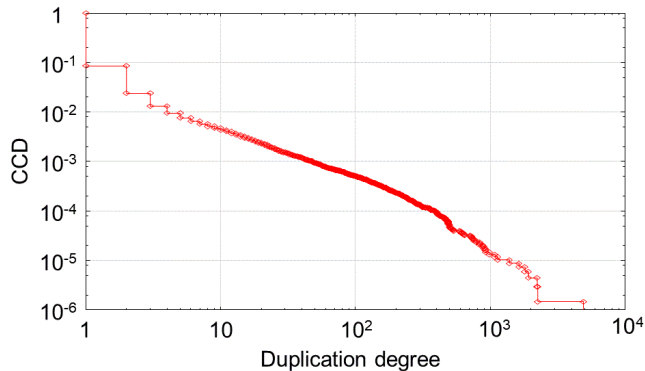


図 4 Complementary cumulative distribution of duplication factor of objects

3.3 オブジェクトの 2 個組の共起度

679,380 個のオブジェクトの任意の 2 個の組の中で、1 つ以上の Web ページで出現した組に対し、8,000 の Web ページにおける共起度 (出現 Web ページ数) の累積補分布を図 5 に示す。約 12 万個の 2 個組が 2 つ以上の Web ページに出現していた。オブジェクト 2 個組の共起度の分布にも冪乗則が観測され、裾野は広く、0.1% 程度の 2 個組は 100 以上の、0.005% 程度の 2 個組は 500 以上の共起度を有する。

次に各オブジェクト組が、特定のカテゴリの Web ページに集中して出現する度合いを調べる。 $n_p(i)$ を、オブジェクトの 2 個組 p が出現したカテゴリ i の Web ページ数と定義し、1 つのカテゴリにおける集中度 f_p を

$$f_p = \frac{\max_{1 \leq i \leq 16} n_p(i)}{\sum_{i=1}^{16} n_p(i)} \quad (1)$$

で定義する。図 6 に各区間 F_j に集中度 f_p が当てはまる 2 個組の数をプロットする。ただし F_j には f_p が $(2j-1)/32 \leq f_p < (2j+1)/32$ の値となった 2 個組が分類される。 j が大きな区間 F_j に分類されたオブジェクト 2 個組ほど、特定のカテゴリに集中して出現していたことを意味する。多数のオブジェクト 2 個組は特定のカテゴリに集中することなく様々なカテゴリの Web ページに広く分散して出現することが確認できる。

3.4 オブジェクトの 3 個組の共起度

次にオブジェクトの任意の 3 個の組の、8,000 の Web ページ

中の共起度を分析する。しかし 1 つ以上の Web ページで出現した全ての 3 個の組数は膨大であり、現実的な時間で共起度を算出することは困難である。よって現実的な時間で結果を得るために、3 個組を構成する各オブジェクトの重複度に下限値 X を定め、重複度が X 以上の任意の 3 個のオブジェクト組に評価対象を限定する。

オブジェクトの重複度の下限値を X とした場合に、全ての構成可能な K 個組の総数に対する、重複度の下限値が X 以上のオブジェクトのみで構成される K 個組 (集合 S_X^K) の総数の割合 (限定率) を R_X^K と定義する。また共起度の下限値 Y を与えたときに、共起度が Y 以上の全ての K 個組に対する、 S_X^K で構成可能な K 個組の割合 (捕捉率) を R_Y^K と定義する。図 7 に 2 個組の限定率 R_X^2 を X に対してプロットする。また図 8 に Y の 5 つの値に対し、2 個組の捕捉率 R_Y^2 を R_X^2 に対してプロットする。

図 8 において、捕捉率は直線 $y = x$ よりも左上の領域に存在し、共起度の評価対象となるオブジェクトの限定率よりも、共起度の高い組の捕捉率が高い。共起度の高いオブジェクト組を構成するオブジェクトは重複度が高い場合が多く、そのため重複度の高い少数のオブジェクトから構成されるオブジェクト組に対してのみ共起度を分析することで、共起度の高い多くのオブジェクト組の共起度の分析が可能である。図 8 において例えば評価する共起度の下限を $Y = 10$ とした場合、全 2 個組の中で約 90% 以上を捕捉できるのは R_X^2 が 0.13 以上の時であるが、図 7 より R_X^2 が 0.13 以上となるのは X が 80 以上の時であった。あくまでも 2 個組の場合であるが、重複度の下限値を 80 に設定することで、全 2 個組の約 90% 以上を捕捉できる。

重複度の下限値 X を 50, 80, 200 に設定した場合の各々について得られた 3 個組集合の中で、共起度が 1 以上の 3 個組の累積補分布を図 9 に示す。図 5 と比較して、 X の増加に伴い捕捉される高共起度組が減少するため共起度の累積分布は右上にシフトする。しかし 2 個組の場合と同様、3 個組の共起度の分布の裾野は広く、例えば $X = 50$ として分析した場合、0.5% 程度の 3 個組は 100 以上の、0.001% 程度の 3 個組は 500 以上の共起度を有することが確認できる。

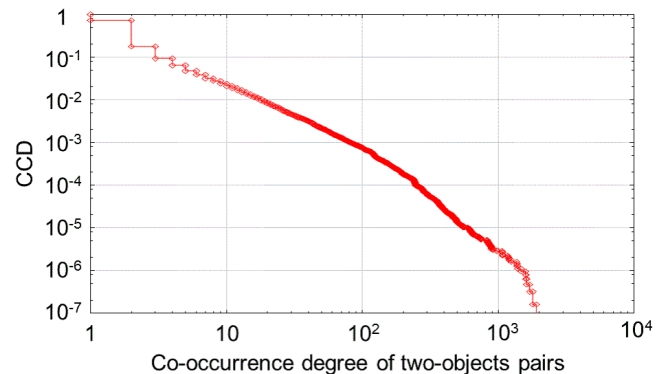


図 5 Complementary cumulative distribution of co-occurrence degree of two-objects pairs

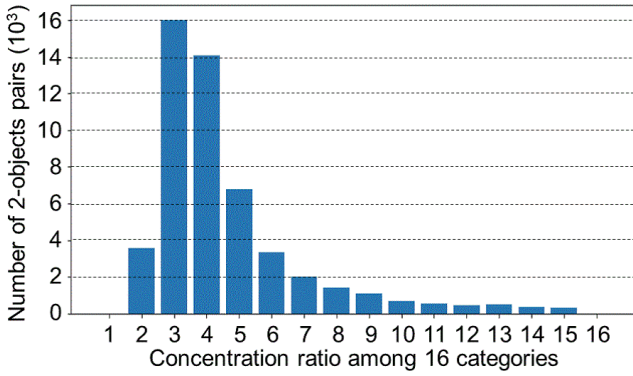


図 6 Concentration ratio of appearance of two-objects pairs among 16 categories

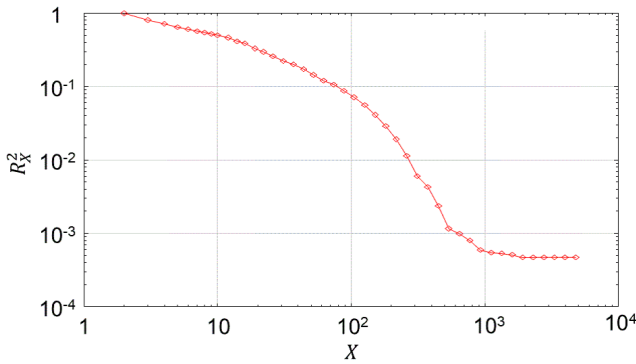


図 7 Restriction ratio of two-objects pairs against lower limit of duplication factor of objects

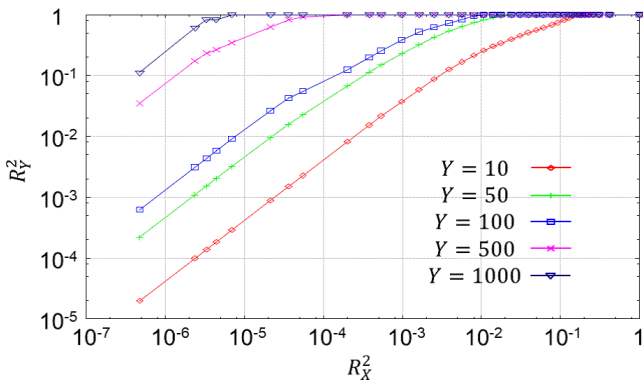


図 8 Capture ratio of two-objects pairs against restriction ratio of two-objects pairs

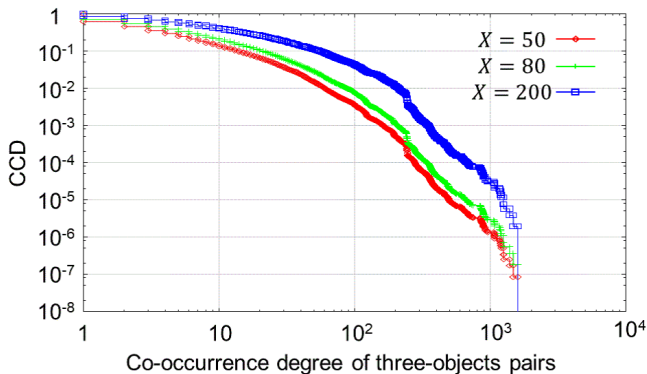


図 9 Complementary cumulative distribution of co-occurrence degree of three-objects pairs

4. 結 論

HTTP/2 の並列配信は同一の配信サーバから取得するオブジェクトに対してのみ可能であり、Web 表示待ち時間の低減効果を高めるには、少数の配信サーバから多数のオブジェクトを取得する必要がある。そこであるオブジェクトの組に対し、その組が出現する Web ページの数を共起度と定義すると、共起度の高いオブジェクト組が優先的にキャッシュに残るようキャッシュ置換を行うことが望ましい。しかし共起度に基づくキャッシュ置換制御の効果は、実際の Web ページにおいてどの程度、オブジェクト間の共起現象が生じるかに依存する。そこで本稿では、共起度に基づくキャッシュ置換制御の可能性を明らかにするため、高人気の 8,000 の Web ページを構成する約 68 万個のオブジェクトを対象に、2 個組と 3 個組の共起度を計算し、Web ページの共起現象の程度を分析した。その結果、共起度の分布は冪乗則に従い、0.1%程度の 2 個組や 0.5%程度の 3 個組は 100 以上の、0.005%程度の 2 個組や 0.001%程度の 3 個組は 500 以上の共起度を有することを明らかにした。そのためこれら少数の共起度の高いオブジェクト組を優先的にキャッシュに残すキャッシュ置換制御の有効性が期待されることを確認した。今後は共起度の基づくキャッシュ置換制御法を検討する予定である。

謝辞 本研究成果は、SCAT 研究費助成 180047 の援助を受けたものである。ここに記して謝意を表す。

文 献

- [1] Alexa, <https://www.alexa.com/siteinfo>
- [2] M. Butkiewicz, H. V. Madhyastha, and V. Sekar, Understanding Website Complexity: Measurements, Metrics, and Implications, ACM IMC 2011.
- [3] When seconds count. <http://www.gomez.com/wp-content/downloads/GomezWebSpeedSurvey.pdf>.
- [4] Not as SPDY as You Thought, Guy's Pod, Thoughts and research on Web Performance & Security, Jun. 2012.
- [5] Software is hard, <http://www.softwareishard.com/blog/har-viewer/>
- [6] M. Jiang, X. Luo, T. Miu, S. Hu, and W. Rao, Are HTTP/2 Servers Ready Yet?, IEEE ICDCS 2017.
- [7] J. Manzoor, I. Drago, and R. Sadre, How HTTP/2 is changing Web traffic and how to detect it, TMA 2017.
- [8] E. Nygren, R. Sitaraman, and J. Sun, The Akamai Network: A Platform for High-Performance Internet Applications, ACM SIGOPS 2010.
- [9] J. Ott, M. Sanchez, J. Rula, and F. Bustamante, Content Delivery and the Natural Evolution of DNS, ACM IMC 2012.
- [10] A. Su, D. Choffnes, A. Kuzmanovic, and F. Bustamante, Drafting Behind Akamai: Inferring Network Conditions Based on CDN Redirections, ACM Trans. Networking, Vol. 17, No. 6, pp. 1752-1765, Dec. 2009.
- [11] S. Sundaresan, N. Feamster, R. Teixeira, and N. Magharei, Characterizing and Mitigating Web Performance Bottlenecks in Broadband Access Networks, ACM IMC 2013.
- [12] X. S. Wang, A. Balasubramanian, A. Krishnamurthy, and D. Wetherall, How speedy is SPDY?, NSDI 2013.
- [13] T. Zimmermann, J. Ruth, B. Wolters, and O. Hohfeld, How HTTP/2 Pushes the Web: An Empirical Study of HTTP/2 Server Push, IFIP Networking 2017.