

ユーザ効用に基づく情報探索の打ち切りに関する一検討

菅原 真司[†] 石橋 豊[†]

[†] 名古屋工業大学 〒466-8555 名古屋市昭和区御器所町

E-mail: †{shinji,ishibasi}@nitech.ac.jp

あらまし 近年、インターネットに代表される大規模ネットワーク上に提供される情報は増加の一途を辿り、ユーザの利便性が飛躍的に向上している。しかしその反面、膨大な情報の中からユーザが真に必要なものを取得することは困難である場合も多く、ネットワークにおける情報探索技術は非常に重要になってきている。インターネットにおける検索エンジンのような探索技術は大きく進歩しているが、一方で、探索候補となるノード（あるいはデータベース、Web サイト）を探索し続けても必要な情報が発見できない場合の対処法については、これまで深く考察されていなかった。筆者らはこの点に着目し、ユーザ効用を考慮した探索打ち切りについての検討を行っている。

キーワード ユーザ効用, 情報探索, 探索打ち切り, 超分散データベース, 大規模ネットワーク

A Study on Utility-Based Truncation for Information Searching

Shinji SUGAWARA[†] and Yutaka ISHIBASHI[†]

[†] Graduate School of Engineering, Nagoya Institute of Technology Gokiso-cho, Showa-ku, Nagoya,
466-8555 Japan

E-mail: †{shinji,ishibasi}@nitech.ac.jp

Abstract As a quantity of information released on a large scale network is getting larger recently, network users' convenience is increased in terms of variety of information. But on the other hand, there are a lot of cases that it is difficult to retrieve the information which the user really wants, and information retrieval techniques are getting more important. Search engines on the Internet have improved dramatically, but we have not investigated what we should do in case we cannot find the information we really want. In this study, we propose a way to truncate the information retrieval on the network, based on utility, and also show one of its effects.

Key words Utility, Information Retrieval, Truncation, Super Distributed Database, Large Scale Network

1. はじめに

近年、インターネットのような大規模ネットワーク上に提供される情報は増加と多様化の一途を辿り、これを利用するユーザの利便性が飛躍的に向上している。しかしその反面、膨大な情報の中からユーザが真に必要なものを取得することは他の情報に紛れてしまうなど困難である場合も多く、ネットワークにおける情報探索技術は非常に重要になってきている。インターネットにおける検索エンジンのような探索技術は日々大きく進歩しているが、一方で、探索候補となるノード（あるいはデータベース、Web サイト）を探索し続けても必要な情報が発見できない場合の対処法についてはこれまで深く考察されていなかった。

筆者らはこれまで大規模ネットワーク内に散在する情報をひとつの巨大なデータベースと捉え、これを超分散データベースと呼び、この中から効率的にユーザの求める情報を取得する手

法について論じてきた [1]~[4]。このような一連の研究の中で、情報探索の打ち切りのための具体的な指針に関する研究も進めてきており、探索打ち切りのための尺度として、統計的手法を用いた探索打ち切りと、探索報酬とペナルティを考慮した探索打ち切りの2つについて既に提案している [5]。本稿ではこの後これらについて簡単に述べる。また、新たな手法として、ユーザ効用を用いた情報探索打ち切りの指針についても述べる。ユーザ効用は時間の関数として、あるサービスを受ける場合のユーザの満足度を表現する。待ち時間が長くなれば一般にユーザの満足度、すなわち効用は単調減少し、ある値に収束する。逆に待ち時間が短かければユーザ効用は大きい。このような時間の関数としてのユーザ効用はこれまでに情報通信ネットワークにおける通信品質の評価にも多数用いられている [6]~[11]。本稿では情報探索の打ち切りの指針としてユーザ効用を用いることを検討し、探索に要するコストを併せて考えることにより、情報探索の適正な継続時間または継続回数を求める手法に関し

て考察を加えている。

2. 情報探索の打ち切りと従来の研究

2.1 情報探索打ち切り問題

大規模ネットワーク上に散在する情報源（情報提供サーバ）が提供する情報の中から、ユーザが必要とする情報（ターゲット）を探索し取得することを考える。このとき本稿では以下の仮定を置く。

- ・ ネットワーク内に散在する探索可能なサーバは N 箇所存在し、 N は十分大きいものとする。
- ・ 各サーバにおけるターゲットの存在確率は既知であり、これらの確率は互いに独立で他のサーバの探索結果により影響を受けない。
- ・ 探索は 1 サーバずつ逐次行い、ターゲットが発見された場合に探索は終了する。
- ・ 探索に要するコストは探索開始時点から探索を終了するまでの総時間（各サーバを探索する場合の時間コストの和）とし、各サーバを探索するための時間は既知である。

以上のような仮定を用いて、これまで筆者らは以下の 2.2 および 2.3 に述べる 2 つの探索打ち切り指針を提案している。

2.2 統計的手法を用いた探索打ち切り

探索対象となる総サーバ数を N 、最大探索回数を r_0 ($\leq N$) とし、この回数までの探索によってターゲットが発見できなかった場合に探索を打ち切るものとする。また、 N 個の各サーバ S_i ($i = 1, \dots, N$) におけるターゲットの存在確率 P_i は既知であると仮定し、簡単のためこれを全て P_0 と置く。この打ち切りによって、探索対象サーバの中にターゲットが存在したにもかかわらず発見できなかった場合を探索エラーと定義すると、この探索エラーが発生する確率 P_f は、

$$\begin{aligned} P_f &= (1 - P_0)^{r_0} P_0 + (1 - P_0)^{r_0+1} P_0 + \dots \\ &\quad \dots + (1 - P_0)^{N-1} P_0 \\ &= (1 - P_0)^{r_0} - (1 - P_0)^N \end{aligned} \quad (1)$$

と表せる。

P_f を r_0 の関数であると考え、 P_f は r_0 の単調減少関数であるから、 r_0 が連続値ならば、ある P_f (ただし $0 \leq P_f \leq 1 - (1 - P_0)^N$) を与える r_0 は一意に決まる。

ここで、統計的な危険率 α を用いて、エラーが発生するという仮説を棄却することを考える。この危険率を一定の値 α に抑制することで打ち切りを行う探索の精度を保証することができる。よって、

$$(1 - P_0)^{r_0} - (1 - P_0)^N < \alpha \quad (2)$$

より、

$$r_0 > \frac{\log\{\alpha + (1 - P_0)^N\}}{\log(1 - P_0)} \quad (3)$$

となるから、自然数の範囲でこれを満たす最小の r_0 を求めることで適正な探索回数を得られる。

しかし、この手法ではターゲットの存在確率が全て一定であ

るという仮定をおかなければならないこと、 α の値と P_0 、 N の値によっては非常に多くの回数の探索を継続することになり、探索エラーの確率を一定値に抑えられても、実際にユーザのニーズに合致しない戦略をとることになる場合が考えられることなどの問題点が残る。

2.3 探索報酬とペナルティを考慮した探索打ち切り

この手法では、これまでに述べた基本的な仮定に以下のものを加える。

- ・ 探索に要するコストは探索時間に相関のあるもののみを考え、ノード i における探索コスト T_i は全てのノードで一定値 T_0 とする
- ・ よって探索は P_i の降順に行われる場合に期待コストが最小となるため、この順に探索を行う
- ・ ターゲット発見に際して得られる報酬を定義し、これを 1 ノードの探索に要するコスト T_0 の定数倍であるとする
- ・ 探索の報酬は探索者およびターゲットにより固有の値であるとする
- ・ 探索を打ち切ることにより探索エラーが生じた場合、探索側にペナルティが与えられるものとし、 T_0 の定数倍であるとする
- ・ 探索エラーによるペナルティは、探索者およびターゲットにより固有の値であるとする

以上の仮定から探索打ち切りの指針を示すために次のような評価関数を考える。ターゲットの発見による報酬を aT_0 、エラーによるペナルティを bT_0 (ただし a, b は非負の定数) とし、 k ノードめ ($k \geq 0$) までの探索でターゲットが発見できなかった場合、 $k+1$ ノードめの探索を行うことにより得られる利得の期待値 E_{k+1} は、

$$\begin{aligned} E_{k+1} &= a \cdot T_0 \cdot P_{k+1} - (k+1) \cdot T_0 \\ &\quad - b \cdot T_0 \left\{ \prod_{i=1}^{k+1} (1 - P_i) - \prod_{i=1}^N (1 - P_i) \right\} \end{aligned} \quad (4)$$

と表せる。このとき、 a, b は探索者およびターゲットにより固有の値をとる。

この評価関数により、探索の事前に E_1 から E_n までの値を計算することで、探索を継続してもターゲットが発見されない場合にいくつめのノードまで探索を行うのが探索者にとって有利であるかを知ることが可能である。

しかしこの手法では、定数 a, b を規定することで探索の報酬とペナルティを定義する必要があるが、何もなところからこの値を仮定することは非常に困難であり、この手法自体に有効性があつたとしても、実際に利用可能なシステムとして実装することは難しい。

3. ユーザ効用と探索打ち切り

3.1 ユーザ効用

ユーザ効用は、ユーザがあるサービスを受けた場合のサービス品質に対する満足度である。探索を継続しているにも拘らず、ターゲット情報が発見できない場合、ユーザ効用は経過時間と

ともに減少する。ターゲット情報とそれを求めるユーザとの関係が探索継続中に変化しないと仮定すると、ユーザ効用は一般に単調減少関数であり、十分長い経過時間の後はある値に収束すると思われる。そこで本稿では、ユーザ効用を以下のように定義する。

- ・ ユーザ効用はユーザとターゲット情報の組み合わせにより決まるものとする。
- ・ ユーザ効用は探索継続時間の関数であり、単調減少関数である。
- ・ ユーザ効用は十分長い時間の後、ある値に収束する。

上記のようなユーザ効用を表す関数としては様々なものが考えられるが、これまでに通信品質の評価に用いる目的で、冪乗関数と指数関数が利用されている [8], [10], [11]。本稿ではこの関数を用いてユーザ効用を表現し、探索打ち切りの基準を定める。

情報探索におけるユーザ効用は、ユーザ個人の特性とターゲット情報の組み合わせにより決まると考えられる。例えばあるユーザ u が、ネットワーク上に存在すると推定されるある情報 I をターゲット情報とした場合、ユーザ効用は u と I との関数となり、さらにこれとは独立に時間 t の関数でもあるとすると、 $U(u, I, t)$ のように表現できる。この関数は前述のユーザ効用の定義を満たす条件下で、ユーザとターゲット情報の組み合わせによりあらゆる形を取りうるが、本稿ではそのユーザの置かれている状況とユーザ視点でのターゲット情報の価値からおおまかに推定できると仮定する。すなわち推定される関数は、ユーザがターゲットを緊急に必要としているならば、ユーザ効用は最初の時点で非常に高く、時間の経過とともに急激に減少する関数となり、逆にある程度時間に余裕があり、比較的長い経過時間の後でもターゲット情報の価値が変化しないならば、緩やかな減少関数となる。ユーザ効用を表現する関数の正確な推定方法は、実際に多くのユーザと多様な探索情報を用いて、学際的な見地も加え、議論されるのが好ましいと思われるが、本稿では検討の対象から除外する。

冪乗関数を U_1 、指数関数を U_2 として表現すると以下の式 (5), (6) のように書ける。ただし簡単のため、ユーザ個人の特性とターゲット情報の組み合わせから決定される要素についてはパラメータ表記から省略し、時間 t のみの関数として扱っている。

$$U_1(t) = At^{-k} \quad (5)$$

$$U_2(t) = Be^{-kt} \quad (6)$$

U_1 における A と U_2 における B は、ともに定数であり、両式における t は探索に要する時間、 k は待ち時間に関する感度に相当する。

さらにこれらの式を用いたユーザ効用関数の典型的な例を以下の図 1, 2 に示す。両図とも、横軸に探索時間、縦軸にユーザ効用を表現しており、各々の図の 3 つのグラフはそれぞれ、待

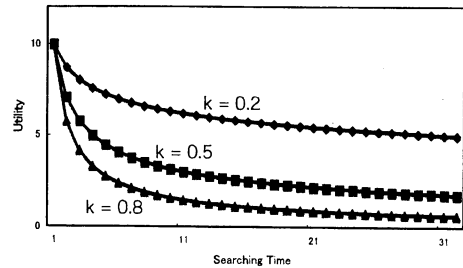


図 1 冪乗関数を用いたユーザ効用の例

Fig. 1 An Example of a Utility Function using Power Function

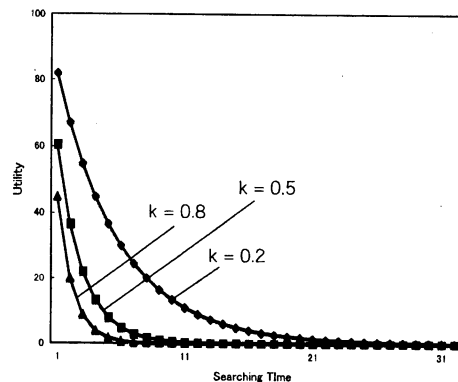


図 2 指数関数を用いたユーザ効用の例

Fig. 2 An Example of a Utility Function using Exponential Function

ち時間の感度を表すパラメータ k の値が 0.2, 0.5, 0.8 の場合の例である。ただし、図 1 では式 (5) における定数 A を 10、同様に図 2 では式 (6) における定数 B を 100 とした。

冪乗関数と指数関数のどちらを用いた場合でも待ち時間の感度 k の値が大きい場合は急激にユーザ効用が減少し、逆に k の値が小さければ緩やかに減少する。また、冪乗関数では一般にユーザ効用は正の値に収束し、感度 k によって収束する値が変化するが、指数関数の場合、ユーザ効用は比較的早い時点でいずれも 0 に収束する。

3.2 ユーザ効用を用いた探索打ち切りの指針

前述のようなユーザ効用関数が推定できる場合、これを用いて探索の打ち切りの指針とすることができると考えられる。ここではその手法について述べる。

大規模なネットワークに接続したユーザが、この中に存在すると考えられる情報（ターゲット情報）の探索を試みた場合、複数の情報源（ノード）からある基準で探索先を選択し、探索を行うことになる。一般に一度に探索できる情報源は複数であると考えられるが、本稿では簡単のため 1 回の探索につき 1 箇所のノードを探索するものとする。（図 3 参照）

最初の探索を行った結果、ターゲット情報を発見した場合に

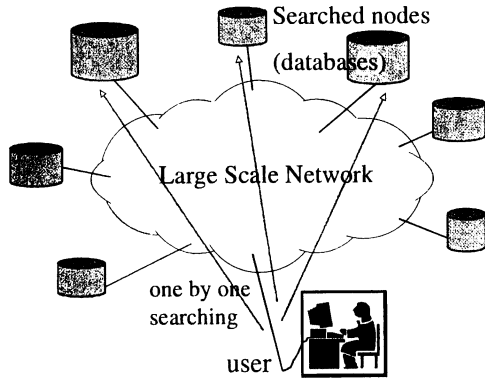


図3 ネットワーク探索の概念図

Fig. 3 A Concept of Information Searching in Large Scale Network

は探索を終了するが、発見できない場合は継続して次の探索先ノードを選択し、探索を行う。各ノードにおけるターゲット情報の存在確率が低い場合、同様の探索行為を繰り返してもターゲットを発見できない事態が発生する。そこで、

- ・ 探索対象となるノードの総数を N とし、探索順に 1 から N までの番号がついているものとする
- ・ ノード i におけるターゲットの存在確率 P_i は互いに独立であり、何らかの方法で推定できると仮定する
- ・ 1 回の探索に要する時間はどのノードの探索でも等しいものとする
- ・ ノード i の探索に要するコスト C_i は既知であるとする

という仮定をおき、 $t > 0$ として、 $t-1$ 回目までの探索に失敗したときの t 回目の探索におけるユーザの利得の期待値 $E(t)$ は以下の式 (7), (8) のように表される。

$$E(t) = \{U(t) - w \cdot C_i\} P_i \quad (t = 1) \quad (7)$$

$$E(t) = \{U(t) - w \cdot C_i\} P_i - w \cdot \sum_{i=1}^{t-1} \left(\prod_{j=1}^i (1 - P_j) \cdot C_i \right) \quad (t > 1) \quad (8)$$

ただし、ユーザ効用を表す関数を一般に $U(t)$ とした。また、探索コストは非負の定数 w による重みを乗じることで、ユーザ効用に換算できるものとした。

$E(t)$ が負であれば探索を行っても利益がないと判断できる。これは時間と探索コストを費やして探索に成功しても、発見した時点では既にその情報に興味を失ってしまうような事態を避けることができ、探索の打ち切りの基準として意味があると考えられる。よって本稿ではこれ以降この手法に従い、探索の打ち切りを行う。

4. 数値解析

前章で述べた探索打ち切りの指針について、計算機による数

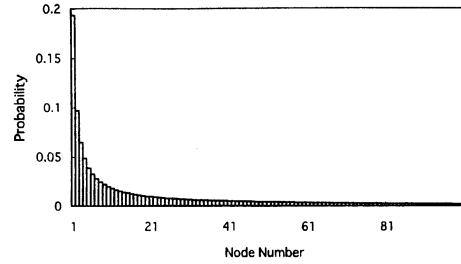


図4 確率密度関数 (1)

Fig. 4 Probability Density Function #1

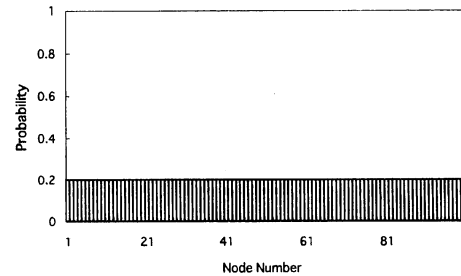


図5 確率密度関数 (2)

Fig. 5 Probability Density Function #2

値解析を行った。ここではさらに以下のようにパラメータ等の条件を定める。

- ・ 探索対象となるノードの総数を 100 とする
- ・ ノード i におけるターゲットの存在確率 P_i の分布を 3 種類用意し、図 4, 5 および 6 に示す確率密度関数と定める
- ・ ユーザ効用の関数は式 (5) を用い、 A の値を 100、 k の値を 0.2 および 0.8 とする
- ・ 1 ノードの探索時間は全て一定値 10 とし、これによりユーザ効用は探索回数の関数として扱う
- ・ ノード i を探索するために要するコスト C_i は 0 から 10 までの一様分布とし、ユーザ効用に換算するための重み w を 0.01 とする
- ・ 上記のパラメータを組み合わせ、それぞれの場合について C_i の分布を 1×10^3 組だけ作り、式 (7), (8) に示したユーザ利得の探索回数ごとの期待値を毎回計算して平均値を求める

以上のような条件で、計算機による数値解析を行った結果を以下の図 8, 9, 10 に示す。これらはターゲットの存在確率、およびユーザとターゲットの組み合わせから決まるユーザ効用に関して典型的なケースとした。また、ここで用いた探索回数の関数となるユーザ効用を図 7 に示す。

図 8 に示した結果は、図 4 の確率密度関数を用い、ユーザ効用のパラメータ値 k は 0.2 とした場合に得られるものである。

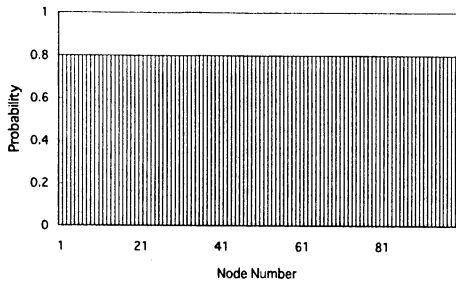


図 6 確率密度関数 (3)

Fig.6 Probability Density Function #3

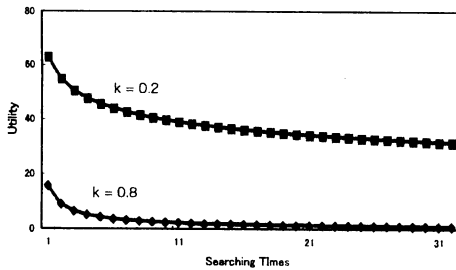


図 7 数値解析で用いたユーザ効用 (探索回数の関数)

Fig.7 Utility Functions used for the Simulation

これを case 1 と呼ぶものとする。case 1 では、各ノードにおけるターゲットの存在確率には大きな偏りがあり、ユーザは比較的時間の経過に影響されずにターゲットを取得したい状況にあると言える。この場合は、ユーザ効用が落ちにくい、比較的長い時間の探索を行うことが可能となっており、12 回目の探索までは利得が正となっている。ユーザはこの程度まで探索を行う価値があるが、それ以降は打ち切るのがよいと考えられる。しかし、ターゲットの存在確率が低いノードも多数あるため、探索の進行に伴い、それらのノードを探索せざるを得なくなると利得値は急落する。また、ユーザ効用の値そのものは高いため、利得の初期値は 12 程度の比較的高い値となる。表 1 に示すように、利得値が正の間だけで探索を打ち切った場合に、最後まで探索を継続すれば発見できたターゲットを見落とす確率を計算すると 0.173 となった。これは高いユーザ効用により、探索回数を重ねる時間的な余裕が生じるため、次に示す、低く、時間的に急落しやすいユーザ効用を用いた場合と比較して小さい値となっている。

図 9 に示した結果は、case 1 と同様に図 4 の確率密度関数を用いているが、ユーザ効用のパラメータ値 k は 0.8 とした場合に得られるものである。これを case 2 と呼ぶものとする。case 2 では各ノードにおけるターゲットの存在確率には case 1 同様に大きな偏りがあるが、ユーザは探索時間の経過に強く影響され、短時間にターゲットが取得できない場合には急速にその興味を失ってしまう状況にあると言える。上記の場合と比較する

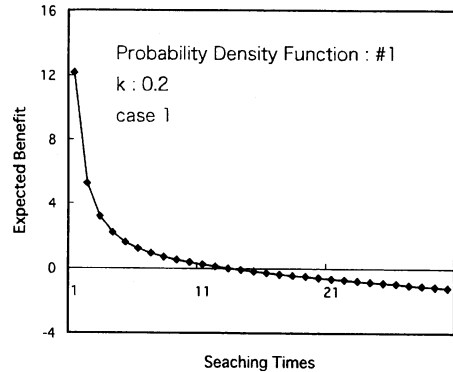


図 8 数値解析結果 (1)

Fig.8 Simulation Result #1

と、利得の初期値は 3 程度、利得値が正である探索回数は 4 といずれも極端に小さくなる。これは低く、急落するユーザ効用の与える影響が顕著に表れていると考えられる。case 2 では 4 回の探索までは行う価値があるが、それ以降は探索を打ち切るのが良いと考えられる。また、探索打ち切りによるターゲット見落としの確率は 0.283 と case 1 よりも大きい。理由は、ユーザ効用の急落により十分な回数の探索をする前に打ち切りを行うためである。

図 10 に示した結果は、ユーザ効用のパラメータ値 k を共通に 0.2 とした上で、図 5 の確率密度関数を用いた場合、および図 6 の確率密度関数を用いた場合の両方をまとめたものである。これらをそれぞれ case 3、case 4 と呼ぶものとする。どちらもユーザは探索時間の経過には強く影響されずにターゲットを取得したい状況にあるが、case 3 ではターゲットの存在確率が全てのノードで比較的低い値の 0.2 としているのに対して、case 4 では全てのノードで極めて高い値の 0.8 としている。これらのケースは特殊であり、探索を長く継続しても利得が落ちにくい。これは、ユーザ効用が十分大きいことに加え、全てのノードで 0.2 または 0.8 のターゲット存在確率があるからである。このような場合は最後まで探索を行う価値があり、その場合には表 1 に示すように、ターゲット見落としの確率は 0 となる。

以上のように、前述の仮定を置くことにより、ユーザ効用を基準とする手法により探索打ち切りの指針を示すことができる。

表 1 探索打ち切りによるターゲット見落としの確率

Table 1 Oversight Error Probability caused by Truncation

case	error probability
case 1	0.173
case 2	0.283
case 3	0.000
case 4	0.000

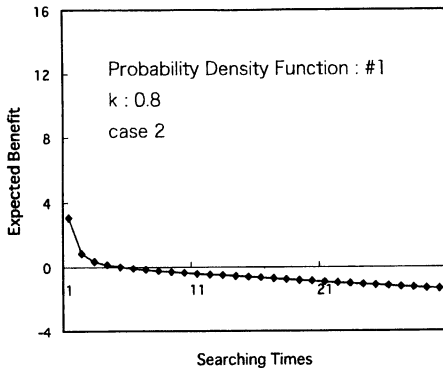


図9 数値解析結果 (2)
Fig.9 Simulation Result #2

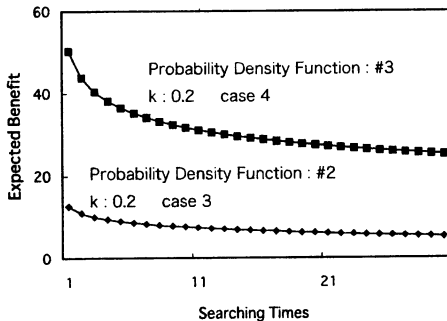


図10 数値解析結果 (3)
Fig.10 Simulation Result #3

5. おわりに

本研究では、多種多様な情報がユーザに公開されている大規模ネットワークにおいて情報探索を行う際に、求める情報が発見できない場合、いつまで探索を継続すべきかという問題について検討を行っている。ここでは探索者が求めている情報を発見した際に感じる満足度をユーザ効用と呼び、探索継続の判断に用いている。ユーザ効用は探索時間の経過とともに単調減少する関数であると定義し、この値から推定される探索利得の期待値が正である期間は探索を継続する価値があると考えられる。

本稿ではこの手法が探索の継続を判断する際のひとつの指針となる可能性を、単純な情報探索の環境を仮定した上で数値解析を行うことにより示した。

今後はさらに探索の条件を具体的なものにすることで、実際のネットワーク内情報探索に貢献できるシステムを検討してゆく。

文 献

- [1] 酒井善則, 山岡克式, 菅原真司, “エントロピーに基づいた情報探索に関する検討,” 信学論 (A), vol.J78-A, no7, pp.894-896,

- July, 1995.
- [2] 菅原真司, ゴーホック-チュー, 大槻一博, 酒井善則, “分散環境における画像情報探索システム,” 画電誌, vol.26, no.4, pp.388-396, April 1997.
- [3] 菅原真司, 山岡克式, 酒井善則, “ネットワークにおける画像情報の効率的探索法に関する検討,” 信学論 B-1, vol.J81-B-1, no.8, pp.484-493, Aug. 1998.
- [4] Shinji Sugawara, Katsunori Yamaoka, Yoshinori Sakai, “A Study on Efficient Information Searches with Agents for Large-Scale Networks”, IEEE Conf. Rec. of Globecom '99, pp.1954-1958, Rio de Janeiro, Brazil, Dec. 1999.
- [5] 菅原真司, 山岡克式, 酒井善則, “超分散情報探索手法に対する探索打ち切りの効果,” 信学技報, IN98-103, CQ98-45, pp.1-6, Oct. 1998.
- [6] 渡辺啓, 滝根哲哉, 村田正幸, 宮原秀夫, “ベストエフォート型サービス/資源予約型サービスにおけるユーザ効用の比較,” 信学論 B, vol.J83-B, no.6, pp.789-799, Jun. 2000.
- [7] 渡辺啓, 村田正幸, 宮原秀夫, “ユーザ効用によるネットワークサービスの比較評価,” 信学論 B, vol.J84-B, no.3, pp.514-522, Mar. 2001.
- [8] 矢守恭子, 田中良明, “優先制御によるデータ転送のユーザ効用最大化,” 信学論 B, vol.J86-B, no.10, pp.2041-2052, Oct. 2003.
- [9] Richard J. La and Venkat Anantharam, “Utility Based Rate Control in the Internet for Elastic Traffic,” IEEE Trans. Networking, vol.10, no.2, pp.272-286, Apr. 2002.
- [10] K.Nomura, K. Yamori, E. Takahashi, T. Miyoshi and Y. Tanaka, “Waiting time versus utility to download images,” 4th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2001), Kathmandu, Nepal/Atami, Japan, pp.128-132, Nov. 2001.
- [11] 梶田剛広, 高橋英士, 矢守恭子, 田中良明, “ユーザ効用に基づくマルチキャスト最適配信間隔設定と呼受付制御,” 信学論 B, vol.J88-B, no.10, pp.2042-2051, Oct. 2005.