

位置情報に基づくオーバレイネットワークにおける 負荷平準化のための推計統計学的手法の提案

Inferential Statistical Approach for Load balancing in Location-based Overlay Networks

小川 和真[†]
Kazumasa Ogawa

中村 陽一[†]
Yoichi Nakamura

齊藤 裕樹[‡]
Hiroki Saito

1. はじめに

近年、移動可能な通信端末の発展により広範囲な位置情報サービスへの期待が高まっている。位置情報を扱ったサービスでは情報量が膨大になるため、P2P ネットワークでサービスを実現する研究が行われている[1]。しかし、従来の位置情報を扱った P2P ネットワークでは、ピアやユーザからのクエリ、センサデータの偏りなどにより特定のノードに負荷が集中してしまう問題がある。

本研究では、オーバレイネットワーク全体の負荷を推計統計学的に平準化するアルゴリズムを提案する。

2. 位置情報に基づくオーバレイネットワークの負荷分散課題

本研究では、位置情報に基づくオーバレイネットワークとして、センサデータを収集する複数のシンクノードで構成されるオーバレイネットワークを想定する。そのようなオーバレイネットワークでは、ピアやユーザからのクエリ、センサデータの偏りなどにより、特定のノードに負荷が集中してしまう可能性がある。ノードの負荷の偏りをなくすために負荷を平準化する必要があるが、位置情報を扱うには範囲検索が必要となるため、キーの順序性を保持しなければならない。また、負荷を平準化するにあたり、自分の負荷がオーバレイネットワーク全体の負荷に比べて小さいのか大きいのかといった負荷の指標が必要となる。

そこで、本研究では、オーバレイネットワーク上の各ノードで推計統計学の中心極限定理を用いて平均と信頼区間を予測し、負荷の指標を作成するアルゴリズムを提案する。また、得られた負荷の指標を用いて Mercury[2]の負荷分散アルゴリズムを基にキーの順序性を保持した負荷平準化アルゴリズムを検討する。

3. 推計統計学を用いた負荷平準化アルゴリズム

3.1 基本設計

オーバレイネットワーク全体の負荷を予測するために、各ノードは中心極限定理を用いて標本平均と信頼区間を算出する。得られた標本平均と信頼区間を用いて各ノードの負荷が信頼区間内になるように平準化する。3.2 節で負荷の指標を作成するアルゴリズムの説明を行い、3.3 節で各ノードの負荷を平準化するアルゴリズムの説明を行う。なお本研究では負荷の定義を以下のようにしている。

負荷 = 蓄積データ数 × 送られてきたクエリ数

[†] 東京電機大学大学院 工学部 情報メディア学専攻
Graduate School of Engineering, Tokyo Denki Univ.

[‡] 東京電機大学 未来科学部 情報メディア学科
Dept. of Info. Systems and Multimedia Design, Tokyo Denki Univ.

3.2 負荷指標作成アルゴリズム

負荷 x の指標となる標本平均の平均と信頼区間を算出するアルゴリズムを図 1 に示す。はじめに、各ノードはランダムサンプリングを 2 回行う。1 回目のランダムサンプリングで識別子と負荷の情報を取得し、負荷の平均を計算する。2 回目のランダムサンプリングで識別子と負荷と標本平均を取得し、標本平均の平均 v を計算する。以降本論文では、標本平均の平均を中心極限平均と呼ぶ。このとき、各ノードはランダムサンプリングしたノードの情報を保持する。次にサンプルサイズを n 、信頼区間の最大信頼区間を p 、最小信頼区間を q とし、式(3)の t は t 分布表に記載されている値を用いて、式(1)で不偏分散 u を計算し、式(2)で標準誤差 s の計算し、式(3)で信頼区間を計算する。

$$u = \frac{\sum_{i=1}^n (x_i - v)^2}{n-1} \quad (1) \quad s = \sqrt{\frac{u}{n}} \quad (2) \quad \begin{cases} p = v + t \cdot s \\ q = v - t \cdot s \end{cases} \quad (3)$$

以下に例を記述する。あるノードがランダムサンプリングして得た標本平均が {19,109,127,52,43,39,67,111,22,51} とする。標本平均の合計が 640 でサンプリングの個数 10 であるから、中心極限平均 v は 64 となる。不偏分散 u は (1)の式より 1324.78 となり、(2)の式より標準誤差 s は 12.13 となる。信頼区間は 95%信頼区間を用いると $t=2.262$ であるから(3)の式より $p=91.44$, $q=36.57$ となる。

Algorithm1. 負荷指標作成アルゴリズム

```
load_barometer begin
while(n < k) do
  randomsampling -> samplingnode;
  n++;
done
sampleaverage = samplingnode.load / k;
while(n < samplesize) do
  randomsampling -> samplingnode;
  n++;
done
//標本平均の平均の計算
limitedaverage = samplingnode.sampleaverage / samplesize
//不偏分散の計算
while(n < samplesize) do
  sum += (samplingnode.load - limitedaverage)^2;
done
unbiased_variance = sum / (m - 1);
//標準誤差の計算
normalerror = (unbiased_variance ÷ samplesize)^1/2;
//信頼区間の計算
reliablemax = limitedaverage + t × normalerror;
reliablemin = limitedaverage - t × normalerror;
end
```

図 1. 負荷指標作成アルゴリズム

3.3 負荷平準化アルゴリズム

得られた平均と信頼区間から負荷を5つの状態に分ける。

1. 最大信頼区間の2倍以上
2. 最大信頼区間より大きいかつ最大信頼区間の2倍未満
3. 信頼区間内
4. 最大信頼区間の半分以上かつ最小信頼区間未満
5. 最大信頼区間の半分未満

以下に、負荷平準化における各状態の動作を示す。

一定間隔ごとに4,5に該当するノードはクラスタリングを行う。4のノードは隣接ノードが4に該当するノードで、ノード数 k 個でクラスタ内合計を c としたとき(4)の式を満たすクラスタを作成し、5のノードは隣接ノードが5に該当するノードで(5)の式を満たすクラスタを作成する。

$$\frac{(k-1)}{c} < p \quad (4) \quad c > q \quad \text{かつ} \quad c < p \quad (5)$$

次に1に該当するノードの負荷を分散する。(6)の式で負荷分散に必要なノード数 m を算出する。ランダムサンプリングした結果から4,5に該当するノードの中から負荷最小のノードを選択し、そのノードのクラスタ内から必要ノードを選択する。4,5に該当するノードがない場合、ランダムサンプリングを行い4,5に該当するノードを探す。選択されたノードは負荷と管理しているキーを隣接ノードに移動し、選択したノードの隣に再参加し、負荷が最小信頼区間以上、最大信頼区間未満になるように負荷とキーを受け取る。

さらに2に該当するノードの負荷を分散する。2に該当するノード最大数 k 個でクラスタリングを行う。隣接ノードが2に該当しないときクラスタを作らない。クラスタ内で負荷が最小のノードは(7)の式で負荷分散に必要なノード数 m を計算する。

$$m = \frac{l}{v} \quad (6) \quad m = \frac{\sum_{i=1}^k (l_i - p)}{v} \quad (7)$$

そして、1のときと同じようにノードを選択し、選択したノードの隣に再参加させる。クラスタ内のノードは最大信頼区間未満になるように負荷を再参加したノードに向けて移動しながら管理キーを変更していく。図2の例はノードID4,8,12のノードでクラスタリングしたノードとノードID64,68でクラスタリングしたノードで負荷分散を行っている。ノードID64のノードはクラスタ内の隣接ノードID68に負荷20とキーの全てを移動し、ID12の隣に再参加する。ID4,8,12は負荷(25,25,10)と対応するキーを再

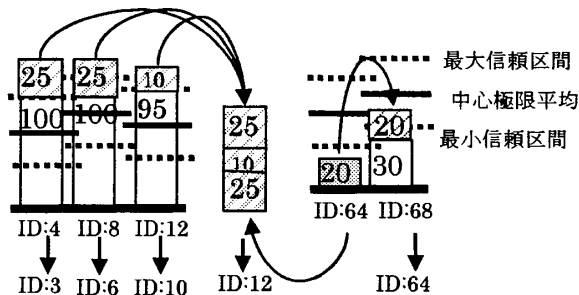


図2. 2に該当するノード群の負荷分散

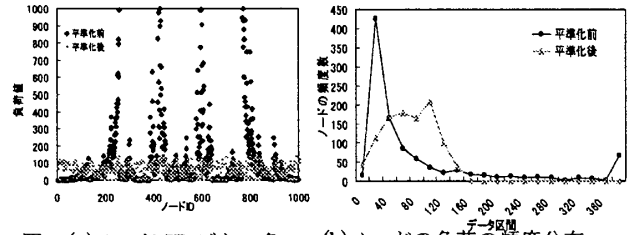


図3.(a)ノードIDごとの負荷の分布 (b)ノードの負荷の頻度分布

参加したノードに向かって移動させ、再参加したノードはIDと負荷を受け取り、ノードID12, 負荷60となる。

4. シミュレーション実験

4.1 評価方法

本シミュレーション実験は、平準化後に何台のノードがオーバレイネットワーク全体の負荷平均に近似するかを評価するために行う。クエリとデータの発生確率は領域の中央を中心とした指数分布になるように行った。シミュレーション条件を以下に示す。

領域の大きさ: $2^{10} \times 2^{10}$, ノード数:1000, 全体のクエリ数:1000, 全体のデータ数:1000, 全体の負荷平均:98, 1回目のランダムサンプリング回数:10回, 2回目のランダムサンプリング回数:20回, 信頼区間は95%信頼区間とする。

4.2 実験結果

図3(a)に各ノードのIDごとの平準化前の負荷の値と平準化後の負荷の値を示す。平準化前と比較すると平準化後は負荷が200以上のノードがなく、50~150に多く分布している。図3(b)に各ノードの負荷の値ごとの頻度分布を示す。平準化前は負荷が20~50と負荷が小さいノードが60%以上かつ少数のノードの負荷が300以上あり偏りが大きいと考えられる。これに対し、平準化後は約65%のノードの負荷が60~120に収まっている。全体の負荷の平均が98で各ノードの最小信頼区間の平均値が63, 最大信頼区間の平均値が124であることから全てのノードの負荷が63~124の間の値になるのが望ましい。シミュレーション結果では、約65%のノードがその区間内に収まった。以上の結果より平準化アルゴリズムが有効であると考えられる。

5. まとめ

本論文では、地理位置情報に基づいたオーバレイネットワークにおける推計統計学を用いた負荷平準化アルゴリズムの提案を行い、評価を行うことで提案手法の有用性を示した。

本研究は、東京電機大学総合研究所研究課題Q07J-02として行ったものである。

参考文献

- [1]金子 雄, 福村 真哉, 春本 要, 下条 真司, 西尾 章次郎: モバイル環境における端末の位置情報に基づくP2Pネットワークの提案と評価, 電子情報通信学会第15回データ工学ワークショップ(DEWS2004)論文集, 2004.
- [2] Ashwin R. Bharambe, Mukesh Agrawal, Srinivasan Seshan: Mercury: Supporting Scalable Multi-Attribute Range Queries, Proc. of ACM SIGCOMM 2004, 2004.