

ネットワーク内に散在する計算資源を用いた 複数情報同時探索のスケジューリング手法に関する研究

学籍番号 23413569 氏名 三浦 力
指導教員名 菅原 真司

1 序論

近年、ネットワーク環境の急速な発展により、インターネット上に非常に多くの情報が蓄積されるようになった。インターネットを利用するユーザは、情報を探索する際に検索エンジンを利用することが一般的となっているが、検索エンジンのシステムでは事前に広範囲にネットワーク内の探索を行った結果を用いるため、更新された直後のコンテンツなどの発見し難いコンテンツが存在する。また、キーワードによる分類が困難であるコンテンツは探索の対象になりにくい。

そこで、グリッドコンピューティングを用いて、ネットワーク上に接続されている多数の計算機の計算能力を利用し、膨大な情報源から直接目的の情報を探索することが考えられる。この場合、ユーザは、柔軟かつリアルタイムの探索が可能になる。

従来研究 [1] では、探索の優先順位の異なる要求情報が発生する場合について十分に検討されていない。ユーザの探索要求には、緊急性を必要とするものと、そうでない探索要求があるため、探索に優先度を付与し、優先度の高い情報が優先度の低い情報より先に発見されるときにユーザの満足度が向上すると考えられる。このことは、ユーザ効用の考え方に合致するものである。本研究では、ユーザが優先度の異なる複数の情報を同時に要求した際に、ユーザ効用が高い状態でこれらを見出し、かつ、ネットワーク負荷を抑制する情報源探索スケジューリング手法を提案し、その評価を行う。

2 情報探索のための資源配分

本研究で仮定している条件を以下に示す。

- 大規模ネットワーク上に、多数の情報源ノード、計算資源ノードが存在する。
- 情報源の管理者は、自身の保有する情報

を公開しており、ネットワークに加入しているすべてのノードから自由に参照できる。

- ユーザは遊休計算資源を情報探索のために利用することができる。

以下では、仮定する問題の定式化とコストの定義を行う。

1. 各情報源の保有するデータ量と各ターゲット情報の存在確率はあらかじめ知ることができ、一般に他と異なるものとする。
2. 情報源の保持するデータは複数の計算資源が分割して探索可能とする。
3. 各計算資源の計算能力はあらかじめ知ることができるものとする。
4. 各情報源の探索では一度に一つのターゲット情報のみを探索可能とする。
5. ユーザは複数のターゲット情報を同時に要求し、それらの探索優先度は一般に異なるものとする。
6. 計算資源が探索対象となる情報源を変更する際に、一定の切り替え時間を要するものとする。

● 時間コスト

探索開始から情報を発見するまでに要した時間とそのターゲット情報の優先度の積を計算し、すべての探索が終了するまでのそれらの合計値を時間コストとする。

● 通信コスト

情報源から計算資源に送信した情報量とその間のホップ数の積を計算し、それらの探索終了までの総量を通信コストとする。

3 キュー制御を用いた計算資源配分スケジューリング

この章では、情報源に計算資源を割り当て、探索のスケジューリングを行う。提案手法では、探索キューに α 個のターゲット情報を挿入し、挿入された数の同時探索を行う。そして、情報源に β 個の計算資源を割り当て、ネットワーク全体の計算資源割り当て数が γ 個に達するまで割り当てを行う。そして、各パラメータを適切に設定することでコストを抑制する。

3.1 評価結果と考察

評価には、第 2 章で説明した時間コストと通信コストを用いる。また、ノード数 10000、情報源数 1000、計算資源数 100 とした。比較手法には、計算資源をクラスタにまとめたうえで探索を行うクラスタアルゴリズム [2]、ランダムな位置の計算資源をクラスタにするランダムクラスタアルゴリズム、全計算資源を用いて探索を行う TOS、最も近い計算資源のみで探索を行う COS の 4 つを用いた。

すべてのコストの値は TOS を用いてターゲット情報を 10 個探索した際に要したコストを 1 として正規化を行った。シミュレーションの結果より、提案手法が最もコストを抑制できることが分かった。また、ターゲット情報が増加するほど、よりコストを抑制できている。

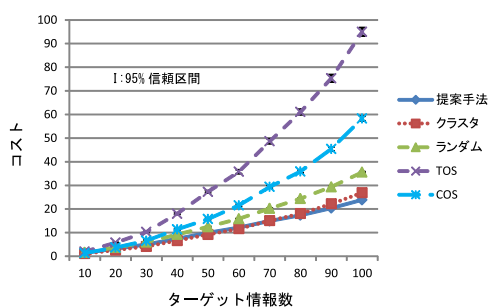


図 1: ターゲット情報数とコストの関係

4 動的パラメータ制御による計算資源配分スケジューリング

この章では、初期状態において、アルゴリズムにおける適切なパラメータ値が未知であるという仮定を置き、効率的なスケジューリング手法を検討する。提案手法では、探索時間や情報源

までのホップ数などを考慮し、情報源に計算資源割り当てを行い、ネットワーク全体の利用数が閾値に達するまで割り当てを繰り返す。この閾値のパラメータ値を動的に変化させることでコストを抑制する。情報の探索を行った際、与えたパラメータで要したのコストを記憶しておき、次の探索には、コストを抑制できたパラメータ値を優先的に選択する。

4.1 評価結果と考察

評価は、3 章の比較手法を用いて、比較を行う。シミュレーション回数は 1000 回とし、平均コストを評価する。図 2 より、提案手法が最もコストを抑制できることが分かる。試行回数は増加するに従い、コストを抑制できる閾値を優先的に選択するため、平均コストが徐々に減少したと考えられる。

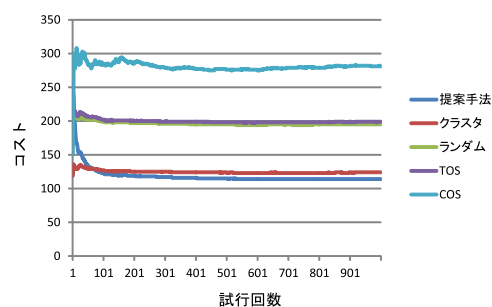


図 2: 平均コスト

5 結論

本研究では、情報探索において、ユーザが優先度の付与された複数のターゲット情報を同時探索する場合を想定して、複数の計算機により直接探索を行う際の効率的な資源配分手法を提案した。また、その評価を計算機シミュレーションにより行った。その結果、今回のシミュレーション環境下では、提案手法が最もコストを抑制できることが分かった。

今後の課題は、断続的にユーザからの探索要求がされる場合など、より現実的なモデルでの検討が考えられる。

参考文献

- [1] 中村, “ネットワーク内遊休資源” 信学技報. NS111(8), 2011.
- [2] 三浦, “ネットワーク内の” 信学技報. NS111(232), 2011.