

エッジコンピューティングにおける物体検出処理を対象とした画像分割転送を用いた負荷分散手法

1240341 合田 和樹

【分散処理 OS 研究室】

1 はじめに

エッジコンピューティングでは、クライアントから距離の近いエッジサーバでクライアントの処理を行い、計算資源への負荷分散や、リアルタイムのデータ処理が可能となる。先行研究 [1] では、クライアント・エッジ・クラウドサーバを対象とした負荷分散手法が提案された。本研究では、先行研究の負荷分散手法に加えて、処理待ちのタスク数の考慮と、物体検出時に負荷に応じて画像を分割転送し、サーバの負荷を分散させる。具体的には、低負荷時には画像を分割せず1台のエッジサーバで画像全体を処理し、高負荷時には、画像を分割し、複数台のエッジサーバで画像を処理を行う手法を提案する。

2 負荷分散手法

2.1 物体検出システムの構成

図1に物体検出システムの構成を示す。本研究では複数のエッジサーバでネットワークを構成し、各サーバがそれぞれ独立しクライアントからのアクセスを、エッジサーバに配置したタスク配置プログラムによって、物体検出タスクをエッジサーバに割り振る。

2.2 提案手法

既存手法では、前回の実行時間と転送時間により、実行時間が最小と推定されたサーバにタスク配置を行う。しかし、データ更新が遅れた場合に、同じサーバにタスクを割り振り、サーバの負荷が増加しレスポンス速度が低下する可能性がある。提案手法では、サーバのレスポンス速度低下を緩和することを目的とし、二段階の項目でタスク配置を行い負荷分散を行う。一段階目では、前回の実行結果と各サーバのタスク実行待ち数を用いて物体検出時間を推定し、その時間が最小となるサーバにタスク配置を行う。二段階目は、物体検出処理自体の処理を緩和するために画像を分割する。推定実行時間が閾値以上である場合に画像を分割し、推定実行時間が二番目に最小となるサーバにタスク配置を行う。

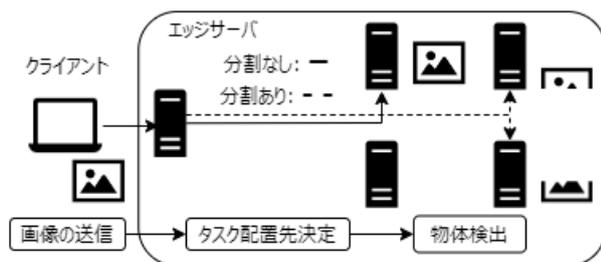


図1 システム構成

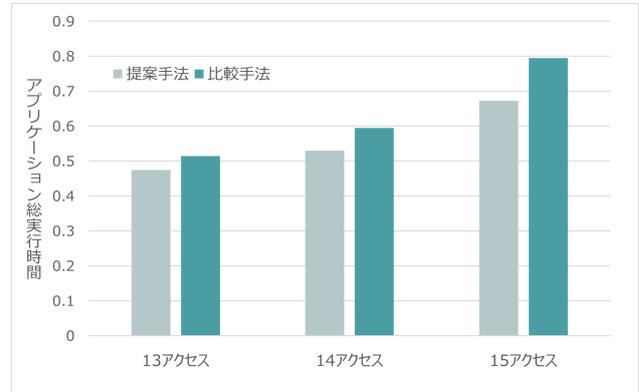


図2 1アクセスごとのアプリケーション総実行時間

3 評価

実験環境としてエッジサーバを5台使いシステムを構築し、閾値を0.7秒とした。性能比較をおこなうために、比較手法として提案手法の画像分割をしないアルゴリズムと比較を行う。また、クライアントは1秒当たり13, 14, 15アクセスを行い負荷をかけた。この状態で得られた結果を図2で示す。比較手法は、アクセスが増加するとアプリケーションの総実行時間が大きくなっている。また、提案手法もアクセス数が増加するとアプリケーションの総実行時間は増加しているが、比較手法と比べてアプリケーションの総実行時間が緩和されていることが確認できた。加えて、物体検出時間は、比較手法に比べて提案手法の方が実行時間が短くなり、アクセス数が増加するほど画像を分割する割合がそれぞれ、71%, 88%, 93%と増加していることが確認できた。これらより物体検出の総実行時間が短くなる理由として、各サーバの負荷が増加した場合に、画像分割を行うことでサーバに配置された物体検出タスクの負荷が減少し、サーバの負荷が減少したためであると推測する。

4 おわりに

本研究では、エッジコンピューティングにおける物体検出処理の負荷分散手法として画像分割転送手法を提案し有効性を評価した。

参考文献

[1] 永元陽一, 横山和俊, "エッジコンピューティング環境での顔認識アプリケーションを対象とした負荷分散機能の実現と評価", 研究報告マルチメディア通信と分散処理 Vol.2021-DPS-187, No12(2021).5