

アクセス予測を用いたビデオサーバの負荷分散制御

3V-7

高田 智規 深澤 勝彦 佐々木 主税

NTT ヒューマンインタフェース研究所

1 はじめに

分散型ビデオサーバでは、特定のサーバやコンテンツに対する集中的なアクセスがシステム全体の性能劣化を引き起こすという問題がある。これを避けるためには、サーバ間で負荷を分散することが必要である。

本稿では、ニューラルネットワークを用いてアクセス数の予測を行い、予測値を用いた負荷分散制御を行う方法を提案し、シミュレーションによってその有効性を評価する。

2 ビデオサーバの負荷分散法

2.1 分散型ビデオサーバの問題

複数のサーバを高速ネットワークで接続した分散型ビデオサーバでは、コンテンツ流通の容易化・コスト削減・拡張性向上が可能であるが、特定のサーバやコンテンツへのアクセス集中が起きた時にシステム全体の性能が劣化してしまう。そのため、分散型システムにおいてはサーバ間で負荷を偏り無く分散させる制御（負荷分散制御）を行う必要がある。

負荷分散制御として、端末からの視聴要求時に同じコンテンツを蓄積しているサーバの中から最適なサーバを選択するアクセス制御と、処理負荷の高いサーバから低いサーバへのコンテンツコピーがある。しかし、コンテンツコピーにはそのコンテンツの視聴時間に比例したコピー時間が必要であるため、サーバの負荷が大きくなってからコピーを行ったのでは、コピーが完了する以前にサーバが処理限界に達しアクセス拒否が起こる。そこで、十分なだけ事前にサーバの負荷を予測し、その予測結果を基にコンテンツコピーを行う事でアクセス拒否を減少させることを考える [1]。

2.2 負荷分散方法

分散型ビデオサーバの各サーバは現在・過去のアクセス数などの環境情報を取得し、未来の負荷（アクセス数）を予測する。また、予測したアクセス数を基に、制御選択（コンテンツコピーの有無、コピー先サーバの選択）を行う。

端末からのアクセス要求を受けた場合にはアクセス制御を行う。これは端末の要求したコンテンツが蓄積されているサーバのアクセス数を調べ、最もアクセス数の少ないサーバに対して接続を試みる、というものである。

また、一定間隔でサーバのアクセス数から未来のアクセス数を予測する。この予測アクセス数が閾値を超えた場合にコンテンツコピーを行う。コピーするコンテンツ

はその時点で最も人気のある（アクセスの多い）コンテンツであり、最もアクセスの少ないサーバへコピー可能であれば（そのコンテンツが存在せず、負荷が閾値を超えていなければ）そのサーバをコピー先サーバとする。

2.3 アクセス予測方式

各サーバはアクセス数の予測を行うが、アクセス特性は時間帯・コンテンツの種類や数などによって動的に変化していく。システム設計時にこのようなアクセス特性を把握し、アクセス数を予測する関数を与えることは非常に困難である。そこで、学習によって動的なアクセス特性に追従しアクセス予測を行うことを考える。

このような学習に対する要求条件は、1) 時間的に連続したデータの学習、2) 実時間での処理、の2点であり、これらの要求を満たすことができるリカレントニューラルネットワーク（以下RNN）を採用する [2]。また、時刻ごとの平均アクセス数を入力に用い複数周期先の予測が可能な学習方法 [3] によって予測精度を向上させることとする。

3 負荷分散のシミュレーション

RNNによるアクセス予測を用いた負荷分散制御方式の有効性を検証するためにシミュレーションを行った。

3.1 シミュレーション

表1の構成でシミュレーションを行った。

端末はあらかじめ与えられたアクセスデータファイルに基づいて、あるコンテンツへのアクセス要求を出す。サーバのアクセス数が既に最大であればアクセスが拒否され、そうでなければコンテンツの視聴時間の間、サーバが提供されサーバのアクセス数は1増加する。ここで、順逆早送り・スキップ再生などの特殊再生は考えない。また、アクセスデータファイルは文献 [4] で用いられたアクセス要求の発生間隔やコンテンツの人気が動的に変化するコンテンツ要求発生モデルによって得られた

表 1: シミュレーション内容

サーバ数	5~50
最大多重度/サーバ	20
最大蓄積数/サーバ	20
コンテンツ数	50
シミュレーション時間	30 日分
コピー閾値	最大多重度の 80%
コンテンツの長さ	50~150 分のランダム
予測時刻	6 時間先
サンプリング・コピー	1 時間おき

値を、各方式で条件を等しくするためにファイルとして保存したものである。

このシミュレーションにおいて、

1. 制御無し (負荷分散を行わない)
2. 予測無し (アクセスが閾値を超えればコピー開始)
3. 提案方式 (RNN の予測値が閾値を超えればコピー開始)

のそれぞれの場合においてアクセス拒否数を測定した。提案方式の RNN には、システム全体に対する 1 日分のアクセスデータを標準化したものを与えて前学習を行った。また、サーバの台数を 5 台~50 台と変化させ、その時のアクセス拒否率 (アクセス拒否数/全アクセス数) の比較を行った。この結果を図 1 に示す。なお、全アクセス数はサーバ台数に比例するよう増減させた。

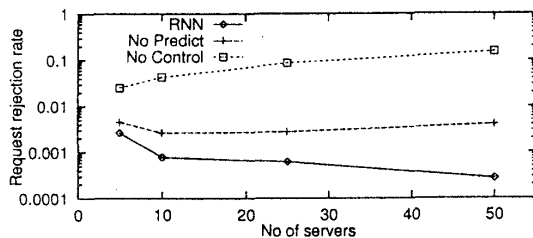


図 1: シミュレーション結果

3.2 考察

比較を行った 3 方式中、提案方式が最もアクセス拒否率が低くなった。また、サーバの台数を増やした時に提案方式以外ではアクセス拒否率が上昇したが、提案方式ではアクセス拒否率が低下した。

提案方式が最良の結果を示した理由として、提案方式では均一な負荷分散が行われているためシステム全体のリソースを効率良く使用することができるのに対し、提案方式以外ではアクセスに偏りが生じ、負荷の高くなったサーバがアクセス拒否を起していることが考えられる。

これを確認するために、サーバが 50 台の時の、各サーバごとのアクセス数とその間の標準偏差を測定した。これを図 2 に示す。横軸は時間、縦軸はサーバのアクセス数を表しており、一番下の値は各サーバのアクセス数間の標準偏差を表している。この図から、提案方式では各サーバ間で均一に負荷を分散させることができていることがわかる。

これらの結果から、本報告で提案した予測・負荷分散方式を用いることで、ビデオサーバの負荷を各サーバに適切に分散させることができ、システムの性能を向上させることができると言える。また、サーバ台数を増やすことで効果的にシステムの拡張も可能である。

4 まとめ

本稿では、RNN によるアクセス数の予測を用いた分散型サーバの負荷分散制御方式の提案、およびシミュレーションを行った。

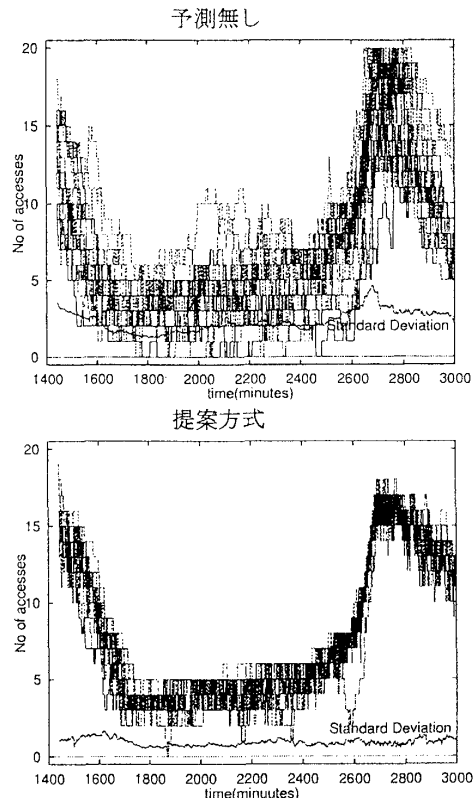


図 2: 各サーバアクセス数・標準偏差

シミュレーション結果から、提案方式を用いることによってビデオサーバの負荷を均一に分散させ、アクセス拒否を低減できることが確認できた。

これにより、提案方式を用いることによって分散型ビデオサーバの性能を向上できると言える。

今後は、RNN 以外の予測方式を用いたシミュレーション、予測精度に応じて動的に予測方式または RNN の構成を変更する機構の検討、NW 通信の遅延を考慮に入れたシミュレーション、実機を用いた性能測定、などを行う予定である。

参考文献

- [1] 深澤, 高田, 佐々木: “分散型ビデオサーバにおける動的負荷分散制御” 第 57 回情報処理学会全国大会 (1998).
- [2] 山下, 高橋, 富田: “簡単なネットワークによる時系列の認識と生成”, 電子情報通信学会技術報告, NC90-116 (1991).
- [3] 高田, 深澤, 佐々木: “分散型ビデオサーバのアクセス予測方式” 電子情報通信学会ソサエティ大会 (1998).
- [4] J. L. Wolf, P. S. Yu and H. Shachnai: “Disk load balancing for video-on-demand systems”, *Multimedia Systems*, 5, pp. 358-370 (1997).