

インターネット活用FB型情報集約技術の評価

—サービス適用とその性能評価—

小泉 泰則 吉田 善幸 長岡 満夫
NTT ソフトウェア研究所

インターネットを利用した視聴者参加型サービスの創出には、双方向コミュニケーションが必須である。しかし、リターンチャンネルは、不特定多数の利用者からのサーバへの情報送信による負荷集中が問題となる。そこで、利用者からの情報受信および情報集約における負荷集中を分散する機能および集約した情報を迅速に利用者に提示する機能を開発している。本稿では、不特定多数の利用者からの情報を集約するサービスにおける利用者からの情報到着状況について時間的変化を分析し、特徴的なパターンを提示する。そして、そのパターンに従って、不特定多数の利用者からの情報集約サービスに適した負荷分散方法について報告する。

Evaluate Method for Collective Feedback Mechanism on the Internet

Yasunori Koizumi, Yoshiyuki Yoshida, Mitsuo Nagaoka
NTT Software Laboratories

When we create collaborative community service for broadcast contents on the internet, it is necessary to communicate each other between server and client. But, return channel of community services has issues in which server processes receive many transactions from many and unspecified users. So we develop the distributed functions of user access and summing up function, and the function of returning the sum to user. In this paper we show the pattern of user information arrival, and focus effective distribution method according to the patterns.

1. はじめに

インターネットを活用して、様々な情報コンテンツに対するリターンチャネルの提供と利用者への集約情報 [投票結果状況、予約状況等] の提供 (これをフィードバック型情報集約方式と呼ぶ) は、視聴者参加型サービス (投票系、予約系、アンケート系等の従来電話系マスコリングサービス) の基盤として重要である。しかし、インターネット利用者の増加に伴い、応答件数がサーバの能力を超え、サービスが十分に機能しないことが想定される。また、インターネットはTV系のマスメディアとの融合も進みつつあり、インターネットを利用した視聴者の応答が特定の時間に集中する確率が高くなるため、従来の集中サービス方式では容易に対処できなくなることが予想される。

本問題に対し、フィードバック型情報集約サービスに適した情報サーバおよびネットワーク負荷の分散、機能の分散による解決方式を開発している (文献1参照) が、実際にサービスを導入する前に受けるであろう負荷を想定し、そのために必要となるマシン数を算出することは難しい。また、ボトルネックとなる部分がCPUか、メモリか、ディスクIOか、ネットワーク負荷かかどうかを探り、サービスに必要な性能を確保することが必要である。しかし、投票パターンは個々の投票が独立したポアソン到着とはならず、投票開始時間、投票締切時間、他の利用者からの投票に誘発された投票等、条件により単位時間当たりの投票数に変化する。このような投票サービス特有の投票数密度 (単位時間当たりの投票数) の時間的変化を分析し、パターン化することにより評価シミュレーションが実態に近づくとともに自動的に投票試験を行い、実態に即したマシン数、ネットワーク性能、マシン性能を確保することが可能となる。本稿では、FB型情報集約サービスにおけるアクセスパターンを分析し、典型的パターンを抽出するとともに、パターンに従って提案するシステムの評価を行う。

2. FB型情報集約サービス

インターネットを使用したフィードバック型情報集約サービスでは、リターン情報はインターネットを使用して送信されるが、送り側はインターネットの他に放送波 (地上波放送、衛星放送、ケーブル放送) で行う方式がある。サービスの情報流通構造を図1に示す。また、サービスの種類としては応答の時間分布の面より(1)企画連動型サービスと(2)企画非連動型サービスに分類できる。(1)はインターネットにおける情報提供者から送信する情報のタイミングと連動して参照しているクライアントの利用者から情報を返送する形態のサービスである。(2)としては送信される情報と連動せずに利用者の意図で情報を返送する形態のサービスである。サービス例を以下に示す。

- インターネットで回答を受けるクイズ番組
- アンケートを収集し、分析結果を視聴者及びサービス提供者に提供する
- リターン情報で購入希望者からの価格を受け取り最高価格を随時提示するオークション
- 娯楽番組で人気投票を受け取り、その結果に従って番組を構成する
- 会議等における案件裁決の投票
- 気象情報、道路情報等の情報を定期的に収集し解析し利用する
- 利用者の情報を参加者で共有することでコミュニティを形成する

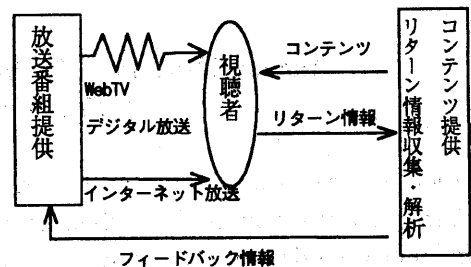


図1 FB情報集約サービスの情報の流れ

3. FB型情報分散集約方式 (WebGONG)

FB型情報集約サービスに対して、利用者からの応答情報の負荷を分散するため以下の機能の特徴とするWebGONGシステムを試作している。

- (1) 情報サーバ・スプリット技術

利用者に各サービスを提供しているコンテンツサーバを意識させることなく、そして、全体の負荷を分散するように利用者の接続先を振り分けるため、企画サーバでは、各サービスを提供しているコンテンツサーバとその提供状況を管理し、利用者からサービスごとに決められた URL を受け取るとフィルタを介して適切なコンテンツサーバを探索する処理を行い、自動的に接続先を切り替える。振り分けは、現状、ラウンドロビン方式のみを適用しているが、負荷制御プロトコルの設計を検討している。また、ネットワーク構成に対応したサーバ配置方式によるネットワーク負荷軽減も実施している。

(2) 分散ギャザリング技術

利用者の個人トランザクション処理は、提供されるサービスの時間的特性により、ある時間に集中する場合がある。そこで、集約負荷の分散を図るため、各コンテンツサーバ毎に一次集約を行い、その結果を集めた全体集約は負荷耐力に従って集約周期を変化させるようにしている。

(3) 集約情報フィードバック技術

全体集約結果を迅速に各コンテンツサーバの利用者に表示するため、一時集約情報収集の応答として各コンテンツサーバに送ることにより、全体集約完了に合わせてコンテンツサーバに表示される。また、集約結果の提供先について、ユーザ/コンテンツ提供者/第3者の事業者など各サービスにより異なるため、可変にできる機能を検討している。

(4) 集約サービス付加技術

利用者からの情報を受け取るサービスでは、不正な情報入力による攻撃が考えられる。そのための個人認証、同一利用者からの多重投票拒否機能等を検討している。また、集約情報の加工編集機能、分析機能について検討をすすめている。

4. 負荷分析

4.1 通常の負荷分析

ユーザのアクセス数負荷を評価するために、アクセス状況をモデル化する方法としては、アクセス間隔をランダムとした指数分布(ポアソン到着)、アクセス間隔を一定とみる単位分布、その2つの分布を一般化したアーラン分布、そして分布の形を制限しない一般分布に分けられる。また、データ発出間隔の周期を複数組み合わせるモデルにより待ち時間の最適化を図る方法等が考えられている。(文献2参照)

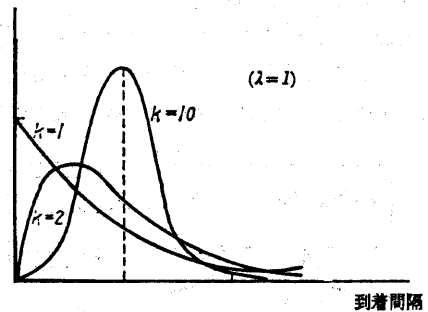


図2 到着分布

4.2 FB型情報集約サービスにおけるフィードバックデータ数のパターン

特定の企画に対応した不特定多数の視聴者からの応答情報については、到着間隔はランダムではなく、サービス開始時や締め切り時、あるいは番組の進行内容により、平均的な到着間隔ではなく特定の時間にピークが集中することや、ある周期でピークが発生するといった状況が考えられる。そのため、これらの集中するときの条件を考慮して、サービスの品質を確保するとともに無駄な設備とならないよう適正な設備設計を行う必要がある。

- ・ 時間遷移による到着間隔・到着情報種別の変化
- ・ 時間遷移によるネットワーク上の通過情報数分布の変化

今回、画面参照数と応答数についての時間的変化を調査し、CPU等の負荷およびボトルネックを分析する。

4. 2. 1 測定方法

FB 型情報集約サービスにおける利用者応答情報の到着状況を分析するため、試作した WebGONG システム上にいくつかの人気投票を受け付けるホームページを設け、インターネット上での投票を受け付け、投票状況を分析することにした。まず、本サービスを行うサーバ上の WWW サーバ機能によりクライアントとの通信内容をログとしてファイルに蓄積した。本通信ログには、通信時の時間、クライアントの IP アドレス、送受信データ等を収録している。その情報より、クライアントから投票画面にアクセスした数とクライアントが投票を行ったときに送信されるデータ名をもつ通信ログ数を抽出し、日別および時間別数を集計した。

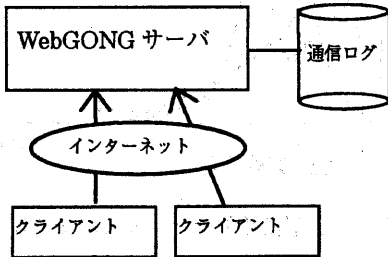


図3 通信ログ収集イメージ

4. 2. 2 測定結果

測定結果として単位時間当たりの投票数と投票画面アクセス数を以下に示す。

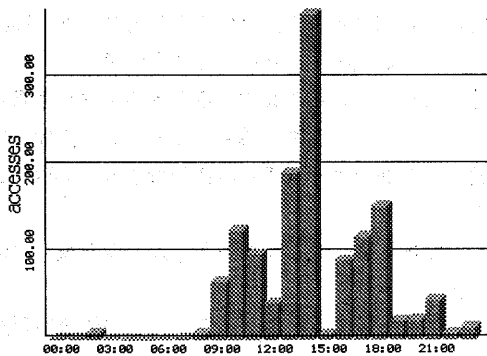


図4 投票数

総投票数	1時間当たり平均投票数
1412	118

表1 投票値

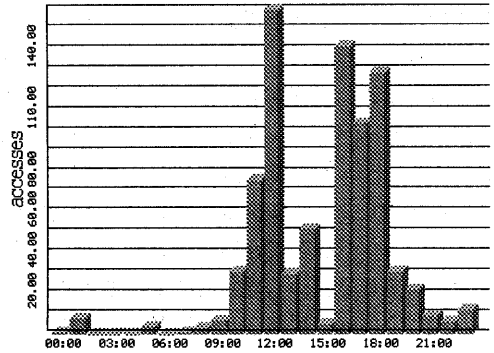


図5 投票画面アクセス数

総画面アクセス数	1時間当たり平均画面アクセス数
834	35

表2 画面アクセス値

4. 2. 3 投票状況の考察

投票数の時間的変化を見ると、ページを参照する数の傾向とはかなり異なっている。まず、投票数が一定しておらず、時間的変動が大きいことがわかる。そして、投票開始時(16時)にしたがって徐々に投票が増加し、投票締め切り時間(翌日15時)前に最大の投票ピークが発生する。夜間帯(22時~翌8時)には投票数が激減している。また、投票数は画面アクセス数に比較してピーク数の集中度が高い。画面アクセス数は12時台が最高値の146件で、総アクセス数が834件のため約18%の集中度に対し、投票数は14時台が最高値の375件で総投票数が1412件のため約27%の集中度である。

WebGONG サービスでは、現時点までの投票集計結果を随時投票画面に提示しているため、投票数の変動が大きいほど投票意志を喚起すること、最終投票結果に投票者の志向を反映させるために投票の最終時期に投票がピークとなると考えられる。したがって、投票に関する負荷については、時間による変動の多さ、ピークに達する間の投票の増加率、ピーク後の激減、ピーク時の投票数と平均値の差を考慮して適正な対応能力の設備を準備することが必要である。

4. 2. 4 投票のパターン

投票結果をもとに以下の3つのパターンに分類し、本パターンによりモデル化することで負荷に耐えられる適正な耐力を分析する。

(1) 駆け込み急増型

投票締め切り前に投票数が増加し、1日の投票数のピークとなる。投票締め切り後は0になる。投票の最終結果を各視聴者が個人の意向に合わせたいため投票数が激増する。

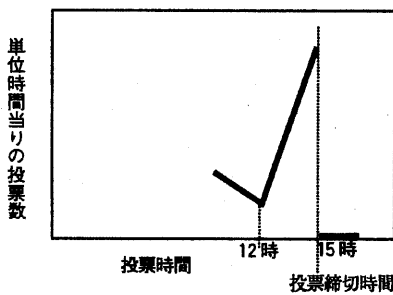


図6 駆け込み急増型グラフ

(2) 開始漸増型

投票開始時のアクセス数の漸次的増加。当初は参照するが投票しない人が多く、徐々に投票数があがってくる。

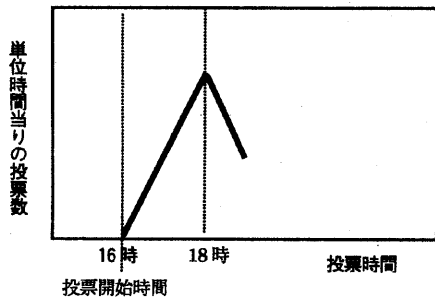


図7 開始漸増型グラフ

(3) 投票者間共振変動型

夜間を除く時間。投票数の増加減少が激しい。投票が行われると、他の視聴者の多くが影響を受けて投票が激増し、投票が減るとまた激減する。投票者間で投票の競争となり激増激減が生ずることも考えられる。

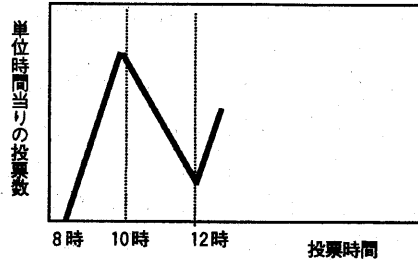


図8 投票者間共振変動型グラフ

5. 性能評価

F B型情報集約サービスの投票アクセスパターンをもとに本システムの評価を行う。

5. 1 性能評価データ

われわれの開発したシステム WebGONG における1投票での投票にかかる性能を示す。

(使用CPU MMXPentium200MHz)

集計	投票
5.3 回/秒	5.0 回/秒

表3 WebGONG 性能値

5. 2 パターンによる評価

(1) 駆け込み急増型による評価

本投票システムでは、1クライアントマシンで毎秒5投票を処理可能である。そのため、最大ピーク投票をXとすると $X/5$ クライアントマシンが必要になる。しかし、ほとんどの時間は無駄となる処理能力を考えると効率的ではない。そこで、ピークの投票を受けつけられる性能として下図のような投票数と投票処理能力に囲まれた斜線の面積分の要求を蓄積できる機能をもたせることで適正な投票処理能力(受けつけた投票のうち単位時間あたりに集計処理可能な数)を設定することが効率的である。

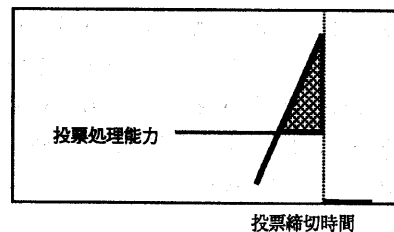


図9 駆け込み急増型評価グラフ

(2) 開始漸増型、投票者間共振変動型による評価

下図のような投票数と投票処理能力に囲まれた斜線のうちの投票処理能力より上位の面積分の要求を蓄積でき、かつ上位の面積が投票処理能力の下位の面積分を超えない条件も満たす範囲で投票処理能力を設定することが効率的である。

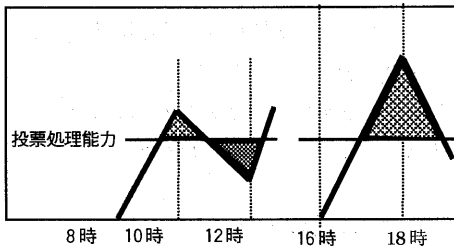


図10 開始漸増型、投票者間共振変動型評価グラフ

6. おわりに

今回、投票サービスについて、その投票数を集計し、投票数の変化をもとに適切な投票処理能力を導出する方法を検討し、以下の結果を得た。

(i)投票サービスのトラフィックは一様/単独の分布には従わず、視聴者間の暗黙のインタラクションに影響され、ページ参照数に比べ変動が激しく、ピークでの集中率が高い。

(ii)投票サービスの負荷パターンは3パターンに大きく分類でき、このパターンに従ってインターネット上の設備の耐応力を決めることが有効である。

(iii) (ii)に沿ってパターン別投票処理能力を、想定投票数の時系列区間による積分比によって決定することが有効である。

今後、より多くのサービスを分析することで普遍的パターンを確立するとともに、集中率やピーク間隔・ピークの時間について数値的に明確にしていく。また、各パターンの原因となる投票者の意識を分析し、影響を与えている投票結果画面の更新間隔、表示内容による投票数の制御についての検討を行う予定である。

参考文献

[1]小泉, ほか: “インターネットにおけるフィードバック型情報集約方式の構想”, 情報処理学会マルチメディアと分散処理研究会, pp31-36, 1997

[2]本間: “待ち行列の理論”, 理工学社, 1973