

平成 26 年度 TSUBAME 産業利用トライアルユース 成果報告書

利用課題名 大容量の行動関連データを使った最適シミュレーション及び暗黙知の利用  
英文: Tacit knowledge and use of optimal simulation using large amounts of data related behavior

佐々木 武  
Takeshi Sasaki

所属  
株式会社 電通国際情報サービス  
Information Service International-Dentsu, Ltd.  
URL <http://www.isid.co.jp>

我々の課題は、膨大なデータが日々蓄積される今日の消費者生活において、広告業務の媒体計画作成に携わる人向けの業務システムを提供しており、顧客が求める期日迄に媒体計画を策定する機能を実現することが困難なケースがあるということである。昨年度、シンプルな課題において高速化が図れたが、より複雑な条件における高速化を目指し、CPU レベルの並列化、GPU レベルの並列化をチューニング検証することが今回の目的である

*Keywords:* Media-Mix Optimized Media Planning Unit-Vehicle

#### 背景と目的

広告キャンペーンにおいて、様々な条件 例えば、「地区」、「媒体種類」、「媒体予算」、「媒体出稿期間」、「最適化指標」などがあり、その条件から媒体への接触ローデータを元に、ビジネスロジックの実装をコンピューターリソースを用いて最適化された媒体計画を算出するアプリケーションを我々は提供している。昨今の広告媒体の状況は取り分け Internet 媒体の種類の高さ、ビジネス・ルールの複雑さなどがめまぐるしい速さで変わっており、媒体組合せが膨大であり、日々増大している。複数の媒体及びその個々メニューのうち、目標に対し最適化された媒体計画を出力するために掛ける時間が膨大になりつつある。集計時間を待つ非効率性と機械損失を含めた

本プロジェクトでは、昨年度の TSUBAME 上での実行時間「4時間58分」に対し半分の時間「2時間30分」を目標に、2倍の処理速度を得るために 既存のボトルネック部分 CPU の並列化を行い、GPU 演算に回る部分も大きな分岐の単位で並列化をめざす。

#### 概要

媒体計画立案のための条件を下記のようにした。

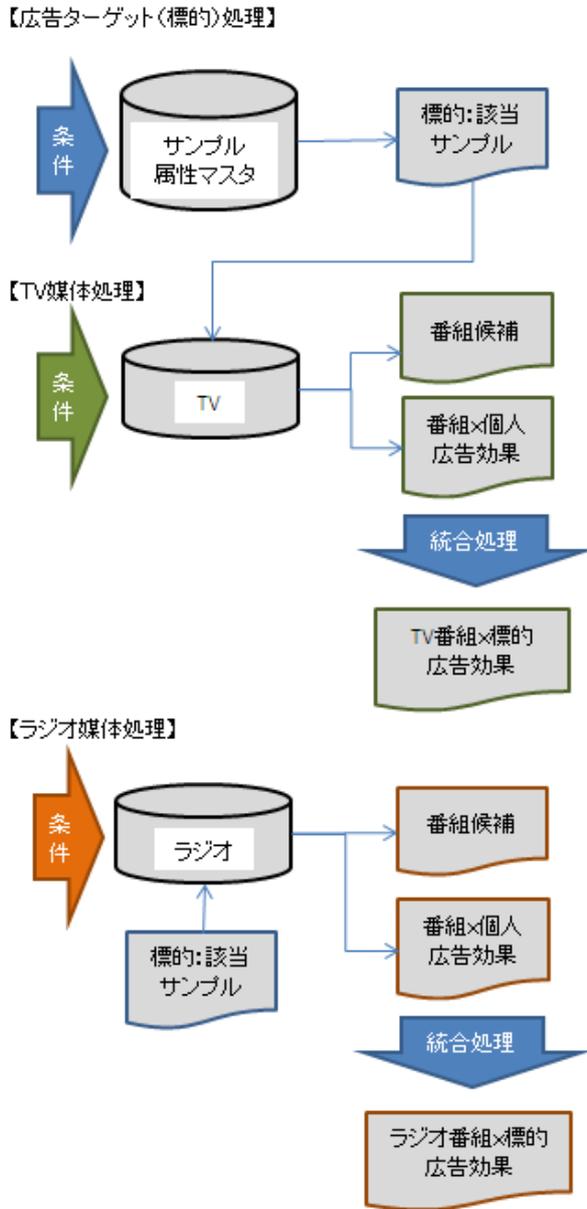
- 広告ターゲット＝「4歳以上男性または女性」
- 広告出稿エリア＝「関東エリア」
- 対象媒体 A 案＝「TV」、「ラジオ」
- 対象媒体 B 案＝「TV」、「ラジオ」、「雑誌」、「新聞」、「屋外・交通」、「Internet」
- 各媒体は スポット広告も番組に見立てて集計
- 出稿予算 A 案＝「1億円」(各媒体 10%刻みで構成)
- 出稿予算 B 案＝「1億円」(各媒体 10%刻みで構成)
- 最適化指標＝広告接触回数「3回」
- 最適化目標＝最適化指標到達率「35%」
- 予算許容指数＝1.2(倍)

これらの条件を元に、最適化された媒体計画を算出する。

結果および考察

◆ 現行の処理概要は下記に処理概要フローを示す。

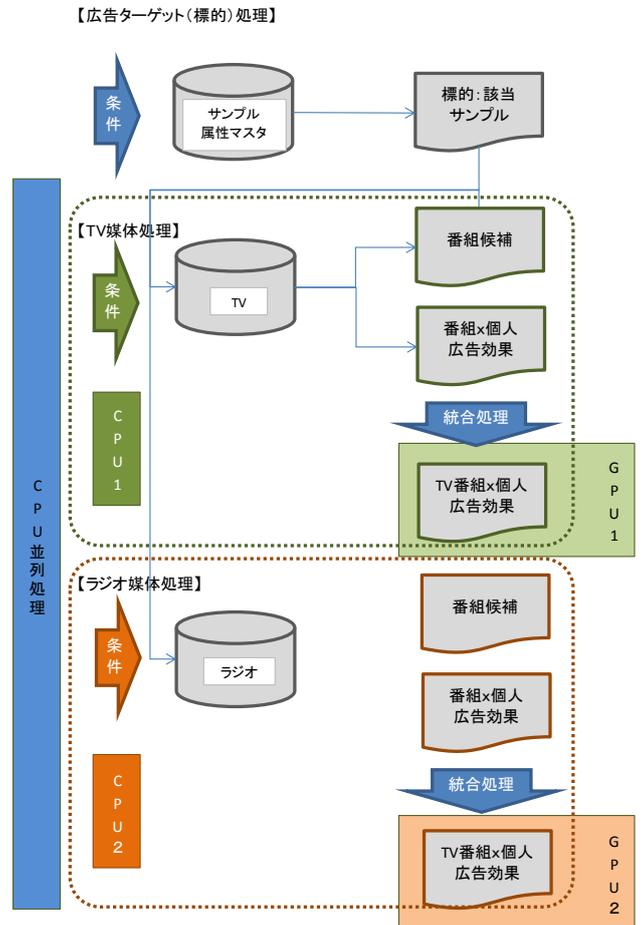
(A 案)



◆ 新たな処理概要は下記に処理概要フローを示す。

(A 案)

- CPU 並列処理導入
- CPU 処理の中に GPU 処理への変更

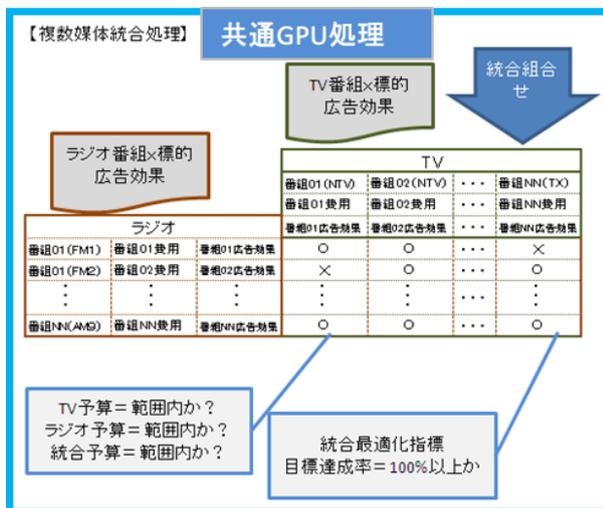


共通の処理は、従来型の GPGPU の処理内容を踏襲。

GPGPU の (B 案) の場合

(A 案) においては、「TV」「ラジオ」の組合せが 2 媒体で固定されていたので、最終的な GPGPU の処理は、1 つの組合せブロックのまま処理するだけで済む。

(B 案) の場合においては、6 媒体の並列処理を行う。この場合、組合せが少ない媒体新聞などは、他媒体より先に終了するが、全媒体の組合せ最適を行うには全部が終了することが必要になる。そのため、管理機能を新たに構築して、「媒体組合せ」、「予算比の組合せ」、「GPU 集計の制御」、「共通部分の GPU 結果共有」、「GPU 結果の保存」の機能を管理機能側が制御し、GPU の演算分と切り分ける。





実測度の計測:

## 1) MPI による CPU 側処理の並列化の影響調査

(同一条件での試行回数 5 回の平均値)

【A 案】2媒体

No.	サーバ様式	経過時間	
1	NormalLinux (NonMPI)	354分	(5時間54分)
2	DualCPU Linux (MPI)	312分	(5時間12分)
3	DualCPU Linux (MPI)【Tuning】	183分	(3時間3分)

CPU 処理の高速化を図り、MPU 処理の昨年比約「1.70 倍」の高速化を実現できた。従来型(一昨年前)の MPI 無しの場合と比較した場合には、「1.93 倍」の高速化が図れた。

【B 案】6媒体

No.	サーバ様式	経過時間	
1	NormalLinux (NonMPI)	675分	(11時間15分)
2	DualCPU Linux (MPI)	596分	(9時間56分)
3	DualCPU Linux (MPI)【Tuning】	348分	(5時間48分)

媒体数を増やした【B 案】の場合においても、今回の Tuning 版・MPU における高速処理の比率は【A 案】とほぼ同等であった。(Tuning 無し)MPI 比「1.71」倍。MPI 無し・比「1.93 倍」という数字であった。

6媒体の組合せにおいて、媒体の特性などがあり新聞や雑誌などは組合せが限定的にしか発生しないため、TV などに比べて合計処理時間が短いため、単純に媒体数の比(2:6)である「3 倍」という時間にはなっていない。(NormalLinux NonMPI で「1.9 倍」)

## 2) GPU 並列化の影響調査

(同一条件での試行回数 5 回の平均値)

【A 案】2媒体

No.	サーバ様式	経過時間	
1	DualCPU Linux + CUDA(GPU)	298分	(4時間58分)
2	Linux (MPI)【Tuning】+ CUDA(GPU)【Tuning】	137分	(2時間17分)

MPI+Tuning による高速化係数 = 「2.17」倍

【B 案】6媒体

No.	サーバ様式	経過時間	
1	DualCPU Linux + CUDA(GPU)	571分	(9時間31分)
2	Linux (MPI)【Tuning】+ CUDA(GPU)【Tuning】	271分	(4時間31分)

MPI+Tuning による高速化係数 = 「2.1」倍

媒体数を増やした「B 案」においても、2媒体の「A 案」とほぼ同じで、MPI+Tuning を行った方が約「2.1」倍高速化が実現できた。ただし、「B 案」の方が「0.07」倍ほど、高速化

の効率が落ちている。この部分は、複数媒体の組合せ集計において「管理機能」が存在するオーバーヘッドが、若干発生している。

まとめ、今後の課題

今期の目情である2媒体で「2時間 30 分」以内は、実現できた。課題については、(B 案)の場合において、6媒体の並列処理で、組合せが少ない媒体新聞などは、他媒体より

先に終了するが、全媒体の組合せ最適を行うには全部が終了することが必要になる。結果一部の GPU のリソースが「待ち」になっている。

これ以上の高速化を図る場合、一番長く処理時間が掛る媒体を更に並列処理をすることが必要である。今回の試みではその部分まで踏むこむことが出来なかった。1媒体あたりの処理を内部で細分化して、共通機構側に進捗情報を共有する仕組みが新たに必要でありこの部分の改善が今後の高速化を図るチューニング・ポイントとなる。