

令和 5 年 5 月 25 日現在

機関番号：14401

研究種目：基盤研究(B)（一般）

研究期間：2019～2022

課題番号：19H02378

研究課題名（和文）大規模データの特徴抽出と再利用に基づくサービス最適割当アルゴリズムの開発

研究課題名（英文）Solving large-scale service assignment problems via machine learning techniques

研究代表者

梅谷 俊治（Umetani, Shunji）

大阪大学・大学院情報科学研究科・教授

研究者番号：80367820

交付決定額（研究期間全体）：（直接経費） 8,000,000円

研究成果の概要（和文）：数百万の利用者を対象とするオンラインサービスにおいて、大規模なデータに基づき利用者の需要に応じたサービスを割り当てる効率的な最適割当アルゴリズムを開発する。特に、多くの入力データが持つ共通した特徴から得られる事前知識を利用し、クエリ時に大規模な入力データ全体を走査することなく最適化計算を実行することで最適割当アルゴリズムの効率化を実現する。

研究成果の学術的意義や社会的意義

現実世界から収集した大規模なデータに基づき利用者の需要に応じて適切なサービスを割り当てる多様なオンラインサービスが現れるようになった。しかし、利用者の需要に応じた商品が推薦されない、需要と供給の不均衡を解消できないなど、適切なサービスが割り当てられない事例が後を絶たないのが現状である。利用者の需要に応じて適切なサービスを割り当てるためには、全ての利用者のクエリを把握した上で、サービスの割り当てが一部の利用者に集中しないように上手く振り分ける全体最適化を実現するアルゴリズムの開発は、社会全体のサービス向上に大いに貢献することが期待できる。

研究成果の概要（英文）：We develop an efficient optimal allocation algorithm for online services with millions of users, which allocates services according to user demand based on large-scale data. In particular, we use prior knowledge obtained from common features of many input data to improve the efficiency of the optimal allocation algorithm by performing the optimization computation without scanning the entire large input data at the time of query.

研究分野：数理最適化

キーワード：組合せ最適化 機械学習 アルゴリズム

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

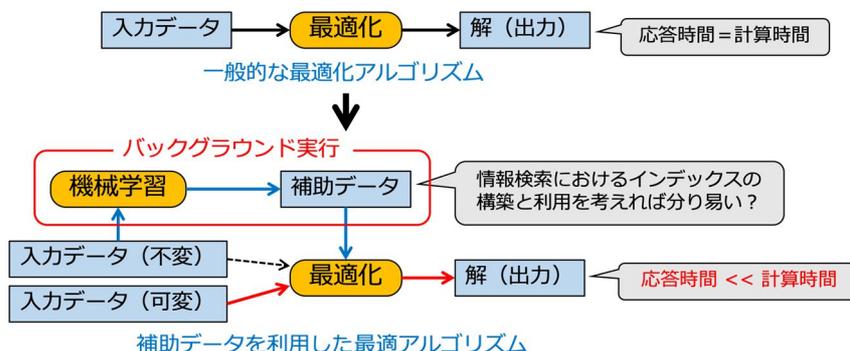
近年、物流・旅客における配車、広告における商品推薦、製造・小売業における受注・発注など、現実世界から収集した大規模なデータに基づき利用者の需要に応じて適切なサービスを割り当てる多様なオンラインサービスが現れるようになった。特に、一般乗用車の運転手と相乗りを希望する利用者をマッチングするライドシェアを始めとして、フリーマーケット、オークション、ジョブマーケットなど、スマートフォンを通じて個人の需要と供給を結び付けるオンラインサービスが急速に拡大している。しかし、多くの利用者がこれらの利便性の高いオンラインサービスを手軽に利用できるようになったにも関わらず、利用者の需要に応じた商品が推薦されない、需要と供給の不均衡を解消できないなど、適切なサービスが割り当てられない事例が後を絶たないのが現状である。

多くのオンラインサービスでは、現場の経験則に基づき、個々の利用者のクエリに対して独立にサービスの割り当てを決定する単純なヒューリスティクスが適用されている（個別最適化）。その結果、サービスの割り当てが一部の利用者に集中し、地域や時刻によりサービスの供給が著しく不足することが少なくない。利用者の需要に応じて適切なサービスを割り当てるためには、全ての利用者のクエリを把握した上で、サービスの割り当てが一部の利用者に集中しないように上手く振り分ける必要がある（全体最適化）。

オンラインサービスにおけるサービスの最適な割り当てを求める問題は、割当問題に個々の事例に応じた実務的な制約条件を追加した整数計画問題に定式化できる。しかし、これらの問題の大半は NP 困難のクラスに属する計算困難な組合せ最適化問題であり、数百万人の利用者を対象とする大規模なオンラインサービスに対して、限られた計算時間で適切なサービスの割り当てを求める効率的な最適割当アルゴリズムを開発することは困難である。

2. 研究の目的

現実世界から収集された大規模なデータでは、新規の入力データが一度だけ与えられるのではなく、多くの共通部分を持つ入力データが繰り返し与えられる状況が多い。そのため、これらの入力データの共通部分からアルゴリズムの性能向上に役立つ特徴を事前知識として繰り返し利用できる事例は少なくない。例えば、文書検索システムでは、検索するキーワードはクエリ毎に変わる一方で、文書データの大幅な変更は頻繁には生じないため、文書データを構造化した補助データを用意することで検索の高速化を実現している。同様に、地図上の経路検索でも、出発地・目的地はクエリ毎に変わる一方で、地図データの大幅な変更は頻繁には生じないため、地図データを縮約した補助データを用意することで探索の高速化を実現している。この方法は、数百万人の利用者を対象とする大規模なオンラインサービスに対して最適割当アルゴリズムを実現する際に、より効果的に働くことが期待できる。本研究では、入力データをクエリ毎に変化する可変な部分とクエリに依存しない不変な部分に分割し、データマイニングや機械学習の手法を用いて、事前に、不変な入力データからアルゴリズムの性能向上に役立つ特徴を抽出して補助データを生成する。補助データの生成・更新と最適化計算を分離した上で、補助データの生成・更新をバックグラウンドで実行し、クエリ時に大規模な入力データ全体を走査することなく補助データを用いて最適化計算を実行することで、大規模なオンラインサービスに対応可能な最適割当アルゴリズムを実現する（下図）。



3. 研究の方法

本研究では、数百万人の利用者を対象とする大規模なオンラインサービスにおいて、大規模なデータに基づき利用者の需要に応じた適切なサービスを割り当てる効率的な最適割当アルゴリズムを開発する。具体的には、急速にサービスを拡大しているオンライン広告に対して効率的な最適割当アルゴリズムを開発する。本課題では、オンラインサービスに現れる大規模な整数計画問題に対して局所探索法に基づく最適割当アルゴリズムを開発する。局所探索法は、現在の解に

対してごく少数の変数の値に変更を加えて(近傍操作)得られる解集合(近傍)の中に改善解があれば,その解に移動する手続きを繰り返す最適化アルゴリズムの基本戦略の一つである.一般に, NP 困難のクラスに属する計算困難な組合せ最適化問題に対して良質の近似解を求めるためには,複数の変数の値を同時に変更する大きな近傍を持つ局所探索法を開発する必要がある.しかし,これらのオンラインサービスに現れる数千万~数十億変数におよぶ超大規模な整数計画問題では,2つの変数を同時に変更する近傍であっても近傍内の解候補の数が非常に多くなるため,限られた計算時間で近傍内から改善解を見つけることは非常に困難である.

本研究では,多くのオンラインサービスに現れる商品推薦最適化問題を考える.この問題は以下の整数計画問題に定式化できる.

【集合】

I: 顧客の添字集合

J: 商品の添字集合

【入力】

$c[i, j]$: 顧客 $i \in I$ に商品 $j \in J$ を推薦するためにかかる費用

$g[i, j]$: 顧客 $i \in I$ に商品 $j \in J$ を推薦したときに得られる利得

C : 顧客に商品を推薦するための総費用の上限

$p[j]$: 商品 $j \in J$ の利得の下限

n : 顧客一人に推薦する商品数

【変数】

$x[i, j]$: 顧客 $i \in I$ に商品 $j \in J$ を推薦するなら 1, そうでなければ 0

$$\begin{aligned} \max. \quad & f(\mathbf{x}) = \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \leq C, \\ & \sum_{i \in I} g_{ij} x_{ij} \geq p_j, \quad j \in J, \\ & \sum_{j \in J} x_{ij} = n, \quad i \in I, \\ & x_{ij} \in \{0, 1\}, \quad i \in I, j \in J \end{aligned}$$

目的関数は総利得の最大化, 1 番目の制約は総費用の上限(予算制約), 2 番目の制約は各商品の利得の下限(利得制約), 3 番目の制約は各顧客に推薦する商品数を表す. 提案する局所探索法では, 実行不可能解も探索するため商品 $j \in J$ の利得制約を緩和し, 対応する重み $w[j] (> 0)$ を用いて, 現在の解の評価関数を以下の通り定義する.

$$z(\mathbf{x}) = \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ij} - \sum_{j \in J} w_j \left[\frac{p_j - \sum_{i \in I} g_{ij} x_{ij}}{p_j} \right]_+$$

ただし, $[a]_+ = \max\{a, 0\}$ である. 近傍内に評価関数の値がより大きな解候補があれば移動する操作を反復する. また, ある一人の顧客 $i \in I$ に推薦されている商品を別の商品に入れ替えて得られる解候補の集合をシフト近傍, ある二人の顧客 $i_1, i_2 \in I$ の間で推薦されている商品を交換して得られる解候補の集合をスワップ近傍と定義する. 局所探索法では, この2つの近傍を交互に探索する.

超大規模な組合せ最適化問題では, 近傍も非常に大きくなるため限られた計算時間内に1回の局所探索法が終了しない. そこで, 近傍から解候補をランダムサンプリングする. しかし, 改善の可能性がない解候補をサンプリングすると探索効率が悪くなるため, シフト近傍, スワップ近傍の特徴を活かした近傍探索を実現する. シフト近傍はスワップ近傍に比べて評価関数の値の変化が大きい. また, 目的関数と利得制約は相反しないため, 一部の商品は多くの顧客に推薦される. そこで, 予算制約を満たす範囲では可能な限り利得および単位費用当たりの利得がより高い商品との入れ替えを考える. スワップ近傍はシフト近傍に比べて解候補の数が非常に多いため, 改善の可能性がない解候補を事前に取り除く. 具体的には, 少なくとも一方の顧客の利得が増加するスワップ操作のみ考慮する.

評価関数のペナルティ重みの初期値は十分に大きな値に設定する. 提案手法では, ペナルティ重みの更新と局所探索法を交互に繰り返し実行する重み付き局所探索法を用いる. 直前の局所探索法で得られた解を \mathbf{x} , これまでに得られた最良解を \mathbf{x}^* とする. 直前の局所探索法で実行可能解が一つも得られなかった場合は, ペナルティ重みが小さすぎると判断し, 以下の通りにペナルティ重みを更新する.

$$w_j \leftarrow w_j + \frac{(z(\mathbf{x}) - f(\mathbf{x}^*)) v_j(\mathbf{x})}{\sum_{j \in J} v_j(\mathbf{x})^2}, \quad j \in J$$

ただし，

$$v_j(\mathbf{x}) = \left[\frac{p_j - \sum_{i \in I} g_{ij} x_{ij}}{p_j} \right]_+$$

である．一方で，直前の局所探索法で実行可能解が一つでも得られた場合は，ペナルティ重みが大きすぎると判断し，以下の通りにペナルティ重みを更新する．

$$w_j \leftarrow \gamma w_j, \quad j \in J$$

ただし， $0 < \gamma < 1$ である．

4．研究成果

提案手法の性能を評価するために，実際の商品推薦の事例を参考に作成した問題に対して数値実験を実施した．顧客数 1 万，10 万，300 万，商品数 500，各顧客に推薦する商品数 5 と設定した問題例に対して数値実験を実施した．また，利得制約の満たし易さの異なる 3 種類の問題例に対して数値実験を実施した．利得制約の満たし易い順に，Type A，Type B，Type C とする．

まず，顧客数 1 万，10 万の問題例について，整数計画問題に対する汎用ソフトウェアである CPLEX12.8，Gurobi7.5 と数値実験の結果を比較した．また，実行可能領域のみを探索する局所探索法の数値実験の結果とも比較した．計算時間は 600 秒で，線形計画緩和問題より得られた上界との相対誤差を評価した．

Type	顧客数	CPLEX12.8	Gurobi7.5	局所探索法 (実行可能領域)	提案手法 高速化あり
A	10,000	34.69%	34.41%	4.06%	0.91%
A	100,000	-	34.38%	7.91%	2.99%
B	10,000	28.87%	0.42%	5.11%	2.01%
B	100,000	-	28.48%	7.68%	3.31%
C	10,000	-	0.53%	3.82%	0.76%
C	100,000	-	17.16%	4.54%	1.81%

汎用ソフトウェア Gurobi7.5 は顧客数 1 万の小規模な問題例で質の良い近似解が得られる場合があるものの，顧客数 10 万の中規模な問題例では質の悪い近似解しか得られていない．一方で，提案手法は，いずれの問題例においても安定して質の良い近似解が得られている．また，実行可能領域のみ探索する局所探索法と比べて質の良い近似解が得られている．

次に，顧客数 1 万，10 万，300 万の問題例について，局所探索法のバリエーションと数値実験の結果を比較した．顧客数 300 万の問題例では計算時間を 3600 秒とした．また，全ての結果で得られた最良解との相対誤差を評価した．

Type	顧客数	局所探索法 (実行可能領域)	提案手法 シフト近傍	提案手法 スワップ近傍	提案手法 高速化なし	提案手法 高速化あり
A	10,000	3.05%	10.99%	30.21%	12.68%	0.00%
A	100,000	5.07%	9.02%	29.56%	4.27%	0.42%
A	3,000,000	9.39%	32.04%	32.61%	8.48%	0.50%
B	10,000	4.14%	10.62%	22.19%	5.39%	0.00%
B	100,000	4.62%	9.46%	21.43%	4.50%	1.00%
B	3,000,000	7.17%	24.18%	24.71%	7.63%	3.31%
C	10,000	3.03%	7.94%	8.61%	2.08%	0.00%
C	100,000	2.05%	6.16%	7.87%	1.61%	0.00%
C	3,000,000	2.93%	10.57%	11.00%	2.79%	0.00%

シフト近傍操作もしくはスワップ近傍操作のみの重み付き局所探索法では質の良い近似解は得られない．高速化なしの提案手法は局所探索法の実行に時間を要し，十分な回数のペナルティ重みが更新できていないため，実行可能領域のみを探索する局所探索法と比べてあまり質の良い近似解が得られていない．一方で，高速化ありの提案手法は十分な回数のペナルティ重みの更新ができており質の良い近似解が得られている．

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Shunji Umetani, Shohei Murakami	4. 巻 303
2. 論文標題 Coordinate descent heuristics for the irregular strip packing problem of rasterized shapes	5. 発行年 2022年
3. 雑誌名 European Journal of Operational Research	6. 最初と最後の頁 1009-1026
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ejor.2022.03.034	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 梅谷俊治	4. 巻 66
2. 論文標題 組合せ最適化による問題解決の実践的なアプローチ	5. 発行年 2021年
3. 雑誌名 オペレーションズ・リサーチ	6. 最初と最後の頁 362-366
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂本 淳子, 大野 修平, 永橋 幸大, 鈴木 保乃加, 梅谷 俊治	4. 巻 66
2. 論文標題 自動車船の運航業務に数理最適化を適用するための実践的なアプローチ	5. 発行年 2021年
3. 雑誌名 オペレーションズ・リサーチ	6. 最初と最後の頁 414-421
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 2件/うち国際学会 0件）

1. 発表者名 濱田賢吾, 棚橋耕太郎, 梅谷俊治
2. 発表標題 ブランドの認知を目的としたTVCM素材の割当て
3. 学会等名 人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 西村直樹, 池田春之介, 木村隆介, 梅谷俊治
2. 発表標題 レベニューマネジメントにおける暗黙知を考慮した最適化モデルの自動構成
3. 学会等名 第23回情報論的学習理論ワークショップ
4. 発表年 2020年

1. 発表者名 神谷俊介, 梅谷俊治, 藤井浩一, 石橋保身
2. 発表標題 整数線形計画問題に対する重みつき局所探索法
3. 学会等名 日本オペレーションズ・リサーチ学会春季研究発表会
4. 発表年 2021年

1. 発表者名 濱田賢吾, 棚橋耕太郎, 梅谷俊治
2. 発表標題 認知視聴者数を考慮したTVCM素材の割当て
3. 学会等名 日本オペレーションズ・リサーチ学会春季研究発表会
4. 発表年 2021年

1. 発表者名 梅谷俊治
2. 発表標題 組合せ最適化による問題解決の実践的なアプローチ
3. 学会等名 第32回RAMP数理最適化シンポジウム(招待講演)
4. 発表年 2020年

1. 発表者名 梅谷俊治
2. 発表標題 組合せ最適化による問題解決の実践的なアプローチ
3. 学会等名 人工知能学会 人工知能基本問題研究会（招待講演）
4. 発表年 2021年

1. 発表者名 中村健吾，藤井浩一，石橋保身，神谷俊介，梅谷俊治
2. 発表標題 集合分割問題に対する重みつき局所探索法の改良
3. 学会等名 京都大学数理解析研究所 研究集会「数理最適化の理論・アルゴリズム・応用」
4. 発表年 2020年

〔図書〕 計1件

1. 著者名 梅谷 俊治	4. 発行年 2020年
2. 出版社 講談社	5. 総ページ数 368
3. 書名 しっかり学ぶ数理最適化 モデルからアルゴリズムまで	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	河原 吉伸 (Kawahara Yoshinobu) (00514796)	大阪大学・大学院情報科学研究科・教授 (14401)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------