A-025

最適モデル多分木による価格需要曲線の推定

伊熊 大貴 [‡] 筑波大学 [‡]
板山 咲穂 [§] 筑波大学 [§]
上原 祐輝 [¶] 筑波大学 [¶]
菊池 明飛 [‖] 筑波大学 [‖]
武井 柊悟 [*] 筑波大学 [*]
竹内 崇貴 [††] 筑波大学 [††]
深谷 悠人 [‡‡] 筑波大学 [‡‡]
柳 智也 [†§] 筑波大学 [†§]
椎名 萌 [‡¶] 筑波大学 [‡¶]
守屋 恵瑠萌 [†‖] 筑波大学 [†‖]
池田 春之介 [‡‖] 筑波大学 [‡‖]
高野 祐一 [†††] 筑波大学 [†††]

1 はじめに

多くの企業にとって，販売商品の価格と需要の関係を理解することは有益である．一般に，商品の価格が上がれば需要は減り，価格が下がれば需要は増える．このような価格と需要の関係を定量的に示す価格需要曲線を得ることができれば，売上や利益を最大化するための商品価格の決定のみならず，ブランディングのための商品の特性の理解や，目玉商品などのプロモーションによる集客施策にも活用可能である．

価格需要曲線の推定を行う際には，解釈性の高い回帰モデルの使用が望ましいと考えられる．解釈性を保ちつつ優れた推定精度を得る回帰分析の手法として，決定木分析がある．代表的な手法として，説明変数の条件で分岐する二分木を構築し，葉ノードにおける事例の目的変数の平均値を予測値とする CART [1] や，葉ノードで任意の機械学習モデルを用いて推定を行うモデル木 [2] などが提案されている．しかし，これらの手法は二分木に基づいているため，予測精度を上げようとすると深い木構造となり解釈性が損なわれる．

この問題に対処するため，整数最適化技術を用いて，一つの分枝ノードから複数のノードへの分岐が可能な多分木を構築する，最適モデル多分木 [3] が提案された．この手法は，モデル木 [2] と同様に葉ノードで任意の機械学習モデルを利用することができ，柔軟なモデリングが可能である．

本研究では，経営科学系研究部会連合協議会主催の令和 5 年度データ解析コンペティションで提供された飲料の購買データを利用する．このデータを元に，解釈性を維持した多分木を構築し，飲料の価格需要曲線の推定を行うことで，価格最適化や集客施策の効果向上に寄与することを目標とする．さらに，最適モデル多分木を構築するための最適化問題の問題規模を縮小して計算を高速化することで，より実用的な手法の開発を目指す．

2 価格需要曲線の推定

本節では，最適モデル多分木を構築する最適化モデルと，その最適化計算の高速化について述べる．

Estimation of price-demand curves using optimal multiway model trees
‡ Daiki Ikuma, University of Tsukuba
§ Saho Itayama, University of Tsukuba
¶ Yuki Uehara, University of Tsukuba
‖ Harutaka Kikuchi, University of Tsukuba
* Shugo Takei, University of Tsukuba
†† Shuki Takeuchi, University of Tsukuba
‡‡ Yuto Fukaya, University of Tsukuba
†§ Tomoya Yanagi, University of Tsukuba
‡¶ Moe Shiina, University of Tsukuba
†‖ Erumo Moriya, University of Tsukuba
‡‖ Shunnosuke Ikeda, University of Tsukuba
††† Yuichi Takano, University of Tsukuba

2.1 最適化モデル

「大分類名：コーヒー飲料，容量：0.4L 未満，販売月：8 月」などの分岐条件の組合せをパスとして定義する．対象とする全てのパスの集合を $\mathcal{P}$ としたとき，最適モデル多分木は $\mathcal{P}$ から全ての事例を包含するようなパスの部分集合を選択することで構築できる．

パス $p \in \mathcal{P}$ に該当する学習事例に対して回帰を行ったときの残差二乗和を ξ_p とする．また，事例集合を $\mathcal{I}$ とし，事例 $i \in \mathcal{I}$ がパス $p \in \mathcal{P}$ に含まれるとき 1，そうでないとき 0 をとる定数を a_{ip} とする．

以上を用いてすべての事例を包含し，残差二乗和を最小にするパス集合を選ぶ最適化問題は以下のような基数制約付き集合分割問題（Set Partition Problem: SPP）として定式化できる．

$$\text{SPP}\left|\begin{array}{lll} \min & \displaystyle\sum_{p\in\mathcal{P}} \xi_p z_p & \\ \text{s.t.} & \displaystyle\sum_{p\in\mathcal{P}} a_{ip} z_p = 1, & \forall i \in \mathcal{I}, \\ & \displaystyle\sum_{p\in\mathcal{P}} z_p \leq \gamma, & \\ & z_p \in \{0,1\}, \quad p \in \mathcal{P}. & \end{array}\right.$$

ここで，z_p はパス $p \in \mathcal{P}$ を選ぶとき 1 を，そうでないとき 0 をとる決定変数であり，1 本目の制約はすべての事例がいずれかのパスに含まれなければならないという制約，2 本目の制約は選ぶパスが γ 個以下でなければならないという制約であり，γ は多分木の葉ノードの数に該当する．目的関数は残差二乗和の最小化である．

2.2 問題サイズ縮小による高速化

定数 a_{ip} から構成される行列を $\boldsymbol{A} := (a_{ip})_{(i,p)\in\mathcal{I}\times\mathcal{P}}$ としたとき，$\boldsymbol{A}$ には重複する行と列が多く含まれる場合がある．実際に数値実験で使用する学習データセットには，同じ事例が多く含まれているため行列 $\boldsymbol{A}$ の行に重複が生じる．したがって，SPP には重複する制約が多数存在する．これらは冗長な制約であるため，削除できる．また，あるパス $p \in \mathcal{P}$ を考えたとき，別のパス $q \in \mathcal{P}$ と包含する事例集合が全く同じである場合がある．その場合には，$\xi_p = \xi_q$ という関係が満たされる．以上より，最適化問題上では，この二つのパス p, q は区別されない．このことより，全く同じ事例集合を含むパスは一つに集約することができる．以上の行・列の集約を行うことで最適性を失わずに計算時間を高速化できる．

また，事例数が少ない葉ノードでの学習は困難であるため，各パスに該当する事例数に閾値を設定し，事例数が閾値未満のパスを削除することで，予測精度を犠牲にせず更なる高速化が可能である．

3 数値実験

本実験では，CART [1]，モデル木 [2]，個別商品の回帰と予測精度，計算時間を比較することで最適モデル多分木 [3] の性能を検証する．実際に使用したデータや前処理の手順，問題設定および結果の詳細は当日報告する．

参考文献

[1] Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14–23.

[2] Quinlan, J. R. (1992, November). Learning with continuous classes. In the 5th Australian Joint Conference on Artificial Intelligence (Vol. 92, pp. 343–348).

[3] Subramanian, S., & Sun, W. (2023, June). Scalable optimal multiway-split decision trees with constraints. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 8, pp. 9891–9899).