

特集にあたって

生田目 崇 (中央大学)

本号は、経営科学系研究部会連合協議会主催「令和4年度データ解析コンペティション」の研究成果について、査読付き論文を募集し、採録された論文により特集を組ませていただきました。7編の論文が投稿され、それぞれの論文について2名に査読をお願いし、結果として4編が採録されました。まずは、短期間で査読をお願いすることになり大変な負担をかけました、査読者の皆様に御礼申し上げます。

令和4年度のコンペティションでは、楽天グループ株式会社様より同社が運営する楽天市場の販売履歴データを提供いただきました。提供いただいたデータの概要は、

- ユーザ情報 (年代 (4 区分), 性別, 居住地域 (9 地域))
- 価格帯 (上位 1 桁で丸め)
- 商品ジャンル (階層構造)

というものであり、データ期間は2019年1月～2020年12月の2年間と、ちょうど新型コロナウイルスの発生前後のデータです。なお、ユーザ数はサンプリングされており、各年代10万人ずつの合計40万人、販売トランザクションはおおよそ1,200万件です。今回のデータの特徴は、何よりも楽天市場で扱っている商品の幅の広さです。小さいものは10円単位から上は億を超える商品まで扱っています (なお高額商品は今回のデータには入っていません)。また、某企業の社長が「使えるポイント上限よりも貯まるポイントの方が多い」という投稿をされていましたが、生活全体で楽天市場を使われている方もいらっしゃる、実際に販売された商品も3,000を超えるカテゴリがあり、今回提供されたデータからどのように購買行動を読み解くかが一つの大きな視点だったと思います。

皆さんも実感がおありと思いますが、新型コロナウイルスの発生直後からしばらくは、この先どうなるのだろうという不安や、学校が休校になり、緊急事態宣言の発出もあり、生活様式や行動の変容を余儀なくさ

れました。そうした中でEC (Electronic Commerce) や宅配での購入が増えたり、また購入するカテゴリも大きく変わったことでしょう。

今回、楽天グループ様には大変な作業をお願いしてデータを提供いただきました。10年くらい前までは、ID付きPOSデータについてFSP番号を変換するなどはするものの、そのまま提供いただくこともありましたが、昨今の個人情報保護法の改正によって、匿名加工情報・仮名加工情報の取り扱いが厳格になったことから、個人情報にあたらないようにデータをさまざま加工してご提供いただく形になりました。今後も (こうした公開のイベントにおいては) 個人が特定できないようなデータしか提供いただけないかなと思っておりますが、制約があるデータからいかにして有用な情報を引き出すか? また、いかにモデルを工夫するか? というという点は同じデータをすべてのチームが共有して行う本コンペティションの醍醐味ということがができるでしょう。今回の採録された論文もORならではのアプローチもさることながらデータの特徴をうまくとらえた分析モデルが各論文で提案されていますので、ぜひご一読ください。

データ分析環境については株式会社NTTデータ数理システム様よりツールを貸与いただきました。毎回快く貸与いただくことに感謝いたします。

令和5年度のコンペティションでは、日本経済新聞社様より、全国の飲料のPOSデータを提供いただいております。コロナ禍も含みながら10年にわたる大変長い期間のデータです。ご興味ある方は3月に開催予定の成果報告会にぜひ足をお運びいただければと思います。令和4年度のコンペティションの成果報告会から対面形式に戻すことができました。コロナ騒動も多少落ち着いてはいますが、街中でのマスク姿などは常にみられるようになり、行動変容はある程度はそのまま維持されているのかと思います。