

2021年度 統計データ分析コンペティション

論文審査会 総評

今回、高校生部、大学生・一般部に共通して、ブラックボックス型AIに基づくチャレンジな論文が寄せられた。AIは、最適な予測モデルを策定するという意味では一定のパフォーマンスを有する一方で、特にブラックボックス型AIの結果を単純な重要度指標で解釈することは危険である。本来定量的な考察に繋げるには計量実証で用いられる説明可能な統計的方法結果を補うなど、緻密な作業が必要である。論文審査会の中でも、これをもって実証研究と言えるかという議論があった。一方で次世代の一つの流れを形成する挑戦的研究活動の芽が息吹いている状況の評価すべきことも確かである。

例年同様、論文審査会では入賞した作品についても様々な問題点があることが指摘されたが、それらは応募論文の主旨を残す意味で必要最小限の修正を除いて手を加えていない。受賞論文の月刊誌「統計」掲載時までには十分解消されるものと期待する。

今回の統計データ分析の過程で得られた新たな知識や経験が、次代を担う若い研究者の糧となることを願って全体総評としたい。

高校生部

コンペティションも4年目を迎え、高校生部は応募論文が大幅に増えており、今回から文部科学省の後援も得て、高校における統計教育が教育現場に浸透してきていることを実感する。応募論文の大半は、高校生らしい統計活用を積み上げつつ、身近な問題に真摯に取り組んでいて、論文構成も大学生・一般部に劣らない読みごたえのあるものも増えてきた。入賞しなかった論文についても、論文審査会は、学校現場の活動を高く評価するものである。

高校生部は全体に論文の質も高くなってきており、データ分析力は以前に比べて格段に上昇してきている。相関分析をベースにした単純で解釈可能な手法に加えて、勾配ブースティングのようなAI技術を活用した意欲的な分析もあった。これについては、全体総評で記載したような相補的教育活動が今後必要と考える。

論文執筆に当たった生徒の方々、あるいはその指導に当たられた先生方に深甚の敬意を表す。

大学生・一般の部

今回の大学生・一般の部は、クラスター分析や機械学習系の論文が上位を占めており、データサイエンスの分野におけるAI関連技術の人気と必要性が高いことを端的に示している。日本の高等教育において、データ分析やプログラミング技術の人材リソースが不足していることは常々感じるところではあるが、若い研究者が先端技術に挑戦する意欲に富んでいることは、高等教育に関わる者として心強い。

統計リテラシーとして、AI関連技術が必須の知識となっていくことは間違いない。一方で、AI分野でも「説明可能なAI」が求められている今日、一部の統計的機械学習に見られるブラックボックス型の分析手法とその出力を基に解釈を行うことのリスクにも留意する必要がある。ブラックボックス型AIの結果をどのように解釈するのか、Sparseモデリングと古典的仮説検定とは何が実証という意味では異なるのかといった、次世代の実証研究のあり方自体を論文審査会でも考えさせられたところである。