データサイエンス教育の題材としての「データ解析コンペティション」

生田目　崇

本誌で毎年 2 月の特集で取り上げられている「データ解析コンペティション」であるが，研究の題材としての利用だけにはとどまらない．むしろ半数以上の参加者は学生（学部生と大学院生）であり，次世代のデータサイエンスを担う若い人材育成の題材として本コンペティションが使われている．本稿では，データ解析コンペティションについて概略を説明した後，このコンペティションの波及効果について研究と教育の両面についてまとめる．さらに，教育面を考慮した場合に，二つの「力」（戦闘力と攻撃力）をキーワードとして，このコンペティションがどのような効果をもたらすのか，また逆に盲点があるのかについて著者の自戒を含めて論じたい．

キーワード：データ解析コンペティション，データサイエンス教育，教育題材

1.　「データ解析コンペティション」について

本誌 Vol. 40, No. 9 の特集「スキャンパネルデータを用いたシェア予測」が本稿で紹介するデータ解析コンペティション（以下，「本コンペティション」）の初回に関する特集記事 [1] である．第 1 回のコンペティションは 1994 年（平成 6 年）に開催されており，さらにその前年に開設された本学会のマーケティング・サイエンス研究部会での活動の一環として始められたわけであるが，それから 26 年間にわたって継続開催しているのが本コンペティションである [2]．おそらく世界的にも最古参のデータ分析に関するコンペティションと思われる．

ちなみに私が参加したのは，第 2 回からであり古参メンバと言えるが，第 1 回から現在まで関わっていただいているのは，後で説明する経営科学系研究部会連合協議会代表の守口剛先生（早稲田大学）と中川慶一郎さん（NTT データ先端技術（株））であり，このお二人を含め先達の方々の活動がすべての始まりである．

本コンペティションについて初めてお知りになられた方もいらっしゃると思うので，最初に簡単に本コンペティションについて紹介させていただく．古くからお知りの方にとっては釈迦に説法かもしれないがお許しいただきたい．

1.1　「データ解析コンペティション」とは

本コンペティションは，各回に企業や各種組織などにご協力いただいて，現場に近いデータ，場合によっては市販されているデータを提供していただき，参加チームが目的を決めそれに合わせた分析を行い発表し合う．データのカテゴリは，もともとが「マーケティング・サイエンス研究部会」から始まったため，マーケティングもしくはその周辺の近い分野のデータがほとんどである．これまでにコンペティションで分析対象としたデータの一覧を表 1 にまとめた．表にあるように基本的には毎年度単一の共通データを提供しているが，複数のデータを提供した年度もある．

コンペティションという名がついており，分析や発表については審査を行い表彰対象チームを決めたりもするが，本来の目的は同じデータを異なる頭脳でさまざまな角度から分析することで，単独の研究では思いもよらなかった知見を誘発・共有することにある．また，シミュレーションデータと異なり，マーケティング分野における実際に消費者が行った行動のデータ（購買や各種ログ）であり，きれいなデータでないことも多く，さまざまなノイズの含まれるデータからいかにして学術的，実務的に有益なメッセージを導くことができるかといった観点も必要となる．

国内外においても kaggle [3] や SIGNATE [4] のような各種のコンペティションがあるが，多くの場合は何らかの予測をし，そのスコアを競っている（ご存じのとおり，高額な賞金まで出ることもある）．本コンペティションはこれらのコンペティションとは異なり，各チームが研究の目的や方向性を決めて，それに合わせて分析を行うためおのずと導かれる結論も異なる．また，提供してきたデータも（発表時には制限がかかる場合もあるものの）商品名がわかる POS データやアクセスログ・データなどの粒度の細かいデータであるため，さまざまなシーンを想像しながら分析目的を考

なまため　たかし
中央大学理工学部経営システム工学科
〒 112–8551　東京都文京区春日 1–13–27
nama@indsys.chuo-u.ac.jp

表 1　年度別テーマとデータ

年度	テーマとデータ	年度	テーマとデータ
H6 年度	食品購買行動	H19 年度	オークション・データ分析
	ストア・スキャン・データ		B2B 自動車オークション・データ
H7 年度	食品・日用品購買行動	H20 年度	消費場面分析
	ストア・スキャン・データ		食卓メニューデータ
H8 年度	日用雑貨品購買行動	H21 年度	百貨店分析
	ホーム・スキャン・データ		百貨店 ID-POS データ
H9 年度	観光行動	H22 年度	日用品 ID 付 POS
	旅行履歴，意識アンケート		ドラッグストア・ID 付き POS データ
H10 年度	食卓マーケティング	H23 年度	ウェブ・マーケティング
	メニュー・データ		ウェブアクセス＋購買履歴
H11 年度	金融マーケティング	H24 年度	サービス・マーケティング
	行動，意識アンケート		クーポン共同購入サイトデータ
H12 年度	金融マーケティング		不動産情報サービスサイトデータ
	銀行取引サマリ	H25 年度	消費者行動分析
	行動，意識アンケート		EC アクセス・購買ログ・データ
H13 年度	流通 CRM		ホーム・スキャン・データ
	ポイントカード・データ	H26 年度	新たな顧客接点
H14 年度	流通 CRM		小売業 FSP データ＋ POS データ
	ポイントカード・データ		ID-POS，EC 購買履歴，アプリ利用
H15 年度	10 周年記念	H27 年度	データの新たな展望
	電力消費データ		複数チェーン ID-POS データ
	ハウスカードデータ		行政窓口受付データ
	スーパー・ドラッグストア POS データ	H28 年度	ファッション EC
H16 年度	2 種類のデータを提供		EC の ID-POS，生活意識調査データ
	加工食品 POS データ	H29 年度	サービス産業の分析
	クレジットカード利用履歴		ヘアサロンの ID-POS データ
H17 年度	アミューズメント POS	H30 年度	生活者のメディア接触分析
	CD 販売店 ID 付 POS データ		メディア視聴・接触データ
H18 年度	ウェブ・マーケティング	R1 年度	都内タクシープローブデータ
	ウェブアクセスログ		タクシーの位置やステータスなどのデータ

えることができる．この点は本コンペティションの大きな特徴といえる．

本コンペティションのもう一つの特徴はすべての参加チームが研究会で口頭発表を行う点にある．上記の他のコンペティションは予測精度が目的であるため，予測値をオンラインで投稿すると精度がすぐに返ってくるというような方式がとられているが，本コンペティションはむしろ着眼点をどこに置くか，というのが出発点であり，分析の目的と結果の整合性，有用性が評価の対象となる．

1.2　開催の体制

本コンペティションは最初の数年は，本学会の研究部会のみで開催してきた．発表チームも 10 チーム程度であり，現在と同様に中間発表と最終発表の 2 回の発表を月に一度開催していた研究会で行っていた．ささやかな規模の研究会であり，参加者みんなで毎回の研究会のあと懇親会をしていたのは懐かしい思い出である．

（以下は，きちんとした記録を残していない部分が多いため，記憶に頼る部分が多いことをお許しいただきたい．）

開催から数年後に，大阪府立大学の 3 名の先生（荒木長照先生，石垣智徳先生（現・南山大学），森田裕之先生）が参加された．徐々に参加者の範囲と人数が増え始めた時期に重なる．また，本コンペティションに参加されていた先生方が，他学会での開催を希望されるようになり，そのとりまとめを前述の中川さんの力添えで（株）NTT データ・技術開発本部（現・技術統括本部）に事務局業務をお願いした．また，現在のように最初の会として「発会式」，最終報告会としての「成果報告会」を開催するようになった．

第 10 回が終わったときに NTT データの事務局業務を一区切することとなり，事務局機能をどうするかという問題が起こった．その解決策として設立した団

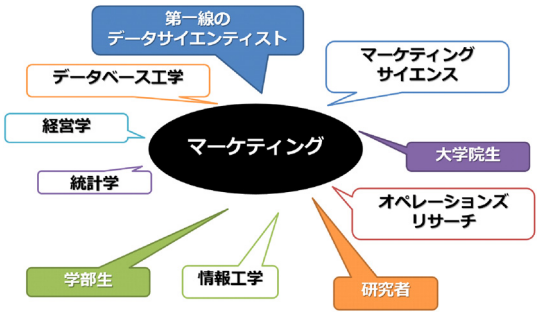

図 1　マーケティングを取り巻く研究領域と人材

体が現在主催している「経営科学系研究部会連合協議会」である．協議会の役割としては，コンペティション開催全体のとりまとめ，データ提供企業との折衝・契約，資金管理などである．なお，コンペティション開催の資金については，固定の資金源はなく基本的には参加チームにご負担いただいている．

当初 7 チームで始めたコンペティションであるが，マーケティング分野が複合領域であることに加え，近年ではビッグデータ，人工知能の主要な分野の一つとして期待されていることもあり近年ではおおよそ 90 から 100 チーム，メンバ総数は 500 から 700 名で開催している（図 1）．開催に協力いただいている学会としては本学会の他には，日本マーケティング・サイエンス学会，日本計算機統計学会，日本データベース学会，日本経営工学会であり，専門領域もマーケティング以外にもハードウェアや統計学，管理技術にまで広くわたっている．また首都圏だけでなく，関西での開催も行っている．所属ベースになるが，日本国内では北は北海道から南は沖縄まで，また，海外の研究者も参加者として登録いただいている．（実際，前年度の成果報告会には現地よりオンライン参加いただいた．大変便利な世の中になったものである．）参加者が増加したため一堂に会しての発表は難しくなったため，各研究部会で発表いただき，選抜されたチームにより成果報告会を開催している．

また，多くのコンペティションが最短で数日，長くても 3 か月程度の開催期間であるのに対して，本コンペティションは 8 月に発会式を行い，各チームの最終発表が 2 月くらいに行われ，さらに成果報告会を 3 月に開催しており，大変長い期間データと向き合うことになる．

この間，協議会の事務局としては，データに関する質問の受付や成果報告会の企画・調整，各種連絡などを行っており，終了後のデータ削除の確認や次節で説明する外部発表の許諾管理などを行っている．さらに，本誌特集号の査読付き論文の募集や論文管理なども本誌編集委員会より委託されている．

2.　研究題材としてのコンペティション

研究に関する議論は本稿の主目的ではないが，簡単に研究材料としてのコンペティションの価値について述べておきたい．

実際の企業からデータを提供いただくということは大変貴重な機会であり，研究者にとっては格好の研究材料である．本コンペティションでも毎年参加いただいている大学教員や研究機関の方々もおり，大変興味深い研究成果をご発表いただいている．各種関連法律に抵触しないようには加工・変数選択をいただいているものの，かなり生データに近いデータを毎回提供いただいており，マーケティングに関する研究だけでなく，消費者行動研究，データ管理，統計学，機械学習・人工知能などさまざまな領域においての適切なデータセットとなっていると思われる．

研究者にとってのもう一つの魅力は，本コンペティションの成果を学会発表や論文投稿など，学術目的に限り本コンペティション終了後（原則 1 年以内に）外部発表を許可していることにあろう．研究成果を外に公表でき，これをきっかけにさまざまな研究交流の加速が期待できる．本誌のコンペティション特集 [1, 5] もその一部であるが，他学会においても研究部会のとりまとめの先生のご尽力により，本コンペティションの成果公開の場をいただいたこともある．また，本誌や JORSJ においても研究成果が投稿論文として掲載された例が複数ある．これも，記憶や主観に頼ることになるが，近年で見ると国内学会や国際会議における研究発表と論文投稿がそれぞれ 20 件程度はある．

近年では提供データも大規模となり（この件については後述する），OR や機械学習，データベース研究において，知識発見手法の提案やアルゴリズム研究の適用事例データとしても使われている．

また，コンペティションの成果はデータ提供企業にお返ししており，研究内容に興味をもってもらえたり，さらなる発展を期待されたような場合は，その後の共同研究や受託研究につながった例もあった．

ただし，データが先にありそれに合わせた分析を行うことを求められるため，事例研究に近い形の研究スタイルになるという弱点もある．しかし，それがゆえの手法の選択やモデリングの工夫，またデータの背後にある本質的な消費者行動を読み解くという研究など

が行えるという意義はあるものと考えている．特に，近年のデータサイエンス研究の高まりとともに，大量データからの有益なルール抽出方法の高度化などが求められており，近年の大規模データの提供はこうした要請にも応えられているのではないかと考えている．

3. 教育題材としてのコンペティション

さて，本稿のメインとなる教育に本コンペティションがどのように貢献できているのか，という点について論じたい．近年の参加者の属性を見ると実はその半数以上を学生（学部生，大学院生）が占めており，場合によっては所属ゼミ生全員に参加いただいているゼミもある．

教育目的の利用としては，次のようなパタンが見られた．

1. 学部生の卒業研究
2. 修士課程学生の研究題材
3. 博士課程学生の博士論文の事例題材
4. ゼミナールにおける討論ネタ
5. 授業の教材

1. から 3. については容易に想像できるかと思うが，実際に卒業・修了に関してコンペティションで提供されるデータを使った学生研究を行う．これらについては，研究の側面もあるだろうが，教員の指導が入るということもあり教育側に入れた．

また，博士論文については，単一の研究成果で博士号を授与されることはないため，コンペティションの成果が博士論文の一つの章などに含まれるという形がほとんどである．またもや記憶に頼ってしまうが，これまでに 10 名程度の博士論文に本コンペティションの成果が含まれている．また，コンペティション参加者がその後大学教員となり研究者人生を歩みながら，後進の指導にあたっている人も複数いらっしゃる．こうした次世代人材を輩出できていることは，私自身一人の大学教員として大変うれしい．

4. と 5. については多少の説明が必要かもしれない．いずれの場合も，構成員（ゼミ生もしくは履修者）全員が参加申込書・誓約書に署名捺印をし，参加している．ゼミの場合はある程度は志向が似通った学生が集まることが想定されるため，そこでの分析もしくは討議の題材として用いられているようである．講義（多くはデータ分析の実習科目）については，同志社大学の宿久洋先生や多摩大学の久保田貴文先生が本コンペティションのデータを取り入れ講義をされている．いずれも，比較的データ分析初学者向けの講義であり，データ分析のイロハから始めて受講者をいくつかのグループに分けたうえで半期の講義の中で何らかのモデル分析を行い，考察を加えて発表するということをしていると聞いている．ご苦労も大変多いかと推察しているが，学生にとっては大変良い経験となっているであろう．

いずれにしても大学教育の場で，実社会の生データに近いデータを実際に利用し，そこからの分析の考察や提案を行えることは，データサイエンス教育にとっても非常に意義深いものと考えている．

また上述したように，本コンペティションでは研究部会での発表を義務付けている．特に学生にとっては，学外の見知らぬ人の前で自分たちの分析の成果を報告し，質疑に答えなければならない．それも，本コンペティションの参加者，すなわち利用しているデータについてよく知っている人たちの前でである．こういった経験は他の発表ではなかなか味わえないものと考えられ，下手な発表はできず，発表準備にも力が入らざるを得ない．それに，他のチームの発表を聞くことで，自分たちでは考え付かなかった分析の方向性や，未知の分析手法についても学ぶことができる．

長年にわたって毎年 2 チーム（これもある種のゼミ内コンペティションなのかもしれない）をエントリーいただき，毎年受賞対象チームとなっている東京工業大学の中田総研 (x)（x は毎年異なる文字）の中田和秀先生が，本誌で取組みについてお書きなのでぜひご一読いただきたい [6]．熱心な指導もさることながら，参加している学生が継続的に時間をかけて分析とディスカッションを行っていらっしゃることで，より良い成果を導き出していることがご理解いただけよう．

大学以外にも，いくつかの企業からの参加は，新人研修の題材になっていると思われるものもある．コンペティションのデータは，実データに限りなく近く，データ分析コンサルティングや各種のシステム開発において，想定されるクライアントに近い場合も多い．もちろん，コンペティションは学術研究の場であるため，データのビジネス利用は許されないが，逆にクライアントのデータでもないためさまざまな試行錯誤ができる．最悪の場合，分析結果からは想定される成果が得られなかったといったこともあろうが，コンペティションの場合は反省と成長の機会にはなっても損害賠償には至らずに済む．

このように，教育の場でこうしたコンペティションに参加するメリットとしては，すでに題材となるデータがあり取り組みやすい点が挙げられる．そのため，どのような方針で分析を進めればよいのか，またその

ためにどのような知識や技術を身に着ければよいのかを効果的・効率的に考えることができる．実は副次的な効果として，参加したということがいろいろなところで学生にとってプラスのポイントになるということもある．場合によっては，奨学金の返済免除の加点対象になったり，就職時に受賞がアピールになったりとのことである（受賞に至らなくともこういう経験を積んだということは評価の対象になろう）．

いずれにせよ，いわば「きれいに整形されていない」データから分析を通じて何らかのメッセージを発信するといったデータ分析の一連の流れを経験でき，その成果を人前で口頭報告する経験を積めることが，本コンペティションの最大の教育効果であると考えている．

4. 期待と自戒

前節までで，これまでの活動について振り返り，教育面からみた効果について自説ながら述べさせていただいた．本節では，これからのデータ活用の潮流や，今一度考え直すべきことについて論じたい．

4.1 今後のデータサイエンス教育とコンペティション

前号ならびに本号の他の記事を読んでいただいてもわかるように，ここ数年，複数の大学でデータサイエンス系の学部・学科の新設や名称変更が続いている．その嚆矢は滋賀大学のデータサイエンス学部であり，別学会での寄稿ではあるがご縁もあってその設立に至るまでのお話を，副学長の須江雅彦先生に論じていただいた [7]．その後，横浜市立大学，武蔵野大学で同学部の設立があり，広島大学情報科学部（コンペティション創設者の木島正明先生が学部長），兵庫県立大学社会情報科学部などでデータサイエンス教育をメイン領域とした学部が設立されている．また，コア分野を名称に入れた一橋大学のソーシャル・データサイエンス学部・研究科，東京医科歯科大学のメディカルデータサイエンス学部の構想など，今後のデータサイエンス領域の拡大が見て取れる．また，東京理科大学のデータサイエンスセンター，早稲田大学のデータ科学センター，本学においても AI・データサイエンスセンターなど他の多くの大学でも大学機関としてのセンターの設立が相次いでいる．学部に限らず，全学教育としてデータサイエンスを進めることは国も推進しており，「数理・データサイエンス・AI 教育の全国展開」の事業大学として 10 の国立大学が選定されており，事業名から見てもその領域の広さがわかる [8]．

大学のみならず，企業においてもデータサイエンスの社内教育は進んでいる．データサイエンティスト協会の設立や，統計士やディープラーニング検定，ウェブ解析士のようなデータ分析や活用に関する資格試験も登場しており注目されている．産学両者におけるさまざまなデータサイエンス分野の進展は今後のこの分野の期待と必要性を示していると言えよう．

4.2 データ解析コンペティションはどこへ行く

さてそのような中で，大して宣伝することもなく毎年 100 チーム近くの参加を得てきている本コンペティションであり，大学をまたいだデータサイエンス分野の積極的な交流チャネルとして活用されてきたと考えており，今後も開催を続ける限りは多くの参加をいただけるものと思っている．反面で，長年続けて開催していることを今一度振り返ってみるといくつかの心配事もある．もちろん，今後も続けて協力いただける企業が見つかるか？ という根源的な心配もあるが，ここではわれわれ教育者が忘れがちな点について触れておく．

少し話が逸れるように思えるかもしれないがここでは二つのドラゴン，「ドラゴンボール」と「ドラゴンクエスト」を取り上げたい．

4.2.1 ドラゴンボールとデータ

ドラゴンボール [9] は鳥山明氏がドクター・スランプに続いて週刊少年ジャンプに連載した大ヒット漫画であり，日本だけで 1 億 6,000 万部，海外を合わせるとその数倍の売上を誇る日本の代表的な漫画である．細かいストーリーには触れないが，10 年にわたる連載の中で，主人公の孫悟空（とその子供たち）は敵と戦い勝利をおさめながら自分の強さを高めつつ，さらに次々と現れる強敵に立ち向かう．その中で「戦闘力」という強さの指標があるので，これを紹介したい．戦闘力はその値が高いほど，高い技術や体力をもち合わせ，戦いの場において有利に行動できる．ドラゴンボールにおいては，最後の方では戦闘力は明示されていないものの，派生して生まれたゲームなどの情報などを含めて，おおよそ表 2 のような推移である[1]．なお，初出話はそのキャラクタが初めて出てきた回を示しており，実際に戦ったシーンではない．

図 2 に横軸 (x) を初出話数，縦軸 (y) を対数変換した戦闘力とした散布図とその近似関数を示す．

近似関数は指数関数を当てはめたが，

$$y = 0.0911 \times \exp\{0.0634x\} \tag{1}$$

[1] 戦闘力の値については諸説あるので，「そうじゃないだろ！」と思う方もご容赦いただきたい．

表 2　ドラゴンボールにおける戦闘力

キャラクタ	初出話	戦闘力
孫悟空（少年時代）	1	10
天津飯	113	180
ピッコロ大魔王	135	260
マジュニア	161	380
ラディッツ	195	416
ベジータ	204	8,000
フリーザ	247	530,000
ギニュー特戦隊	272	120,000
フリーザ最終形フルパワー	321	1 億 5,000 万
セル	361	600 億
魔人ブウ	460	1 兆 2,000 億

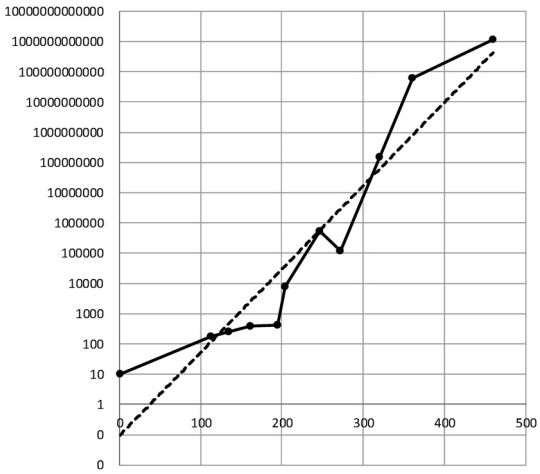

図 2　戦闘力の散布図と近似関数（横軸：話数，縦軸：戦闘力）

となり，決定係数は 0.8981 である．

第 1 回のコンペティションでは，五つのブランドのインスタントコーヒーの購買履歴が提供されたが，データ量は，5,624 行 ×7 列で csv ファイルにするとおよそ 200 kbyte 程度であり，現在から考えると大変ささやかなデータ量とも言える．昨年度（令和元年度）のタクシープローブデータはおよそ 600 Gbyte と，単純計算でおよそ 3,000 万倍となっている．

さて，いったい何が言いたいのかというと，漫画の主人公はいろいろな試練や訓練を通じて自己を鍛えながら強い敵にあたってきたわけであるが，一方で，コンペティションに参加する学生は，コンペティションの場が初めて実データに触れる機会であったりする場合も少なくない．第 1 回コンペティションの時も現在も大学 4 年生は同じ 22 歳であり，同じく第一話で登場した状況に過ぎない．

(1) 式において，$y = 3{,}000$ 万 から逆算すると，

表 3　ドラゴンクエストシリーズと武器の種類

シリーズ	武器の種類	シリーズ	武器の種類
I	7	VII	80
II	15	VIII	121
III	53	IX	265
IV	38	X	375
V	61	XI	258
VI	57		

$x = 309.145$ とおおよそ 310 話程度に相当する．すなわち少年孫悟空（本格的なデータ分析が初めての学部 4 年生）の前に進化を遂げたフリーザ（600 Gbyte のデータ）が突然立ちはだかるようなもので，いきなり「このデータから価値のある分析を行え」といっても多くの場合は途方に暮れるばかりであろう．ICT や IoT の進化とともに，取得可能なデータの粒度は細かくなり次元は拡がった．もちろん，これらの大量のデータからいかにして効果的・効率的に有効な情報抽出をするか，その技術や手腕が問われているのは確かであるが，教育という面で考えると（カリキュラムの変化はあるものの）違わぬ 20 代前半の学生がターゲットであり，脳が処理できる情報量などはデータの増大に比べるとささやかな変化であろう．社会的なデータ分析の要請と育てる人材のレベルについてそのバランスを今一度考えなければならない時期に来ているのかもしれない．

4.2.2　ドラゴンクエストの場合

ドラゴンクエストはエニックス（現・スクウェア・エニックス）が堀井雄二氏の発案をもとに制作したロールプレイングゲームのシリーズである [10]．メインのナンバリングシリーズが 11 作，スピンオフを含めるとさらに多くの作品がある．ナンバリングシリーズは各ナンバーで数百万本の販売実績がある．くしくもキャラクタ・デザインはドラゴンボールと同じく鳥山明氏である．

ドラゴンクエストの世界では敵と戦うための武器をストーリーの進行とともに変えていく．ここでの強さは「攻撃力」である．武器の標準的な攻撃力はドラゴンボールの戦闘力と比べるとあまり大きな変動はなく，第 1 シリーズの最強武器である「ロトのつるぎ」が 40 であるのに対して，最新の第 11 シリーズでは「ひかりの大剣」が 327 と約 8 倍程度である．むしろシリーズを経ての違いは，武器の種類にある（表 3）．

第 1 シリーズにおいてはわずか 7 種類の武器（たけざお，こんぼう，どうのつるぎ，てつのおの，はがねのつるぎ，ほのおのつるぎ，ロトのつるぎ）しかない

のに対して，第 11 シリーズでは現状で 258 種類と格段に増えている．パーティを組むとかそれをさらに入れ替えるといった戦術の自由度が高まったことで，選択可能な攻撃方法が増え，敵に対してさまざまな戦い方ができるようになったし，また適材適所に対応すべきである．

本節の議論は，前節のデータ量の拡大に対するある種の反論であり，本コンペティションを始めた当時とは，コンピュータの能力や分析ツールの性能の向上，新たな分析手法の開発などの，コンペティションに参加するうえで，武装できる武器（ツールや技術）に格段に進化があり，学生であっても比較的容易にこれらを利用できることは確かであり，いかにツールを駆使して，やりたいモデリングを実装できるというのも研究手段の一つではある．本コンペティションの開始当時ではそもそも利用可能なコンピュータでこれらのデータを読み込むことすら難しかった．今や AWS や GCP などのクラウドサービスなどを使えば高価なハードウェアを用意することなく，巨大なデータを扱える環境がすぐさま用意できる．

したがって，前節のデータ量のみでの比較は公平ではないかもしれない．分析手法においても当初は多変量解析系の分析がメインストリームであったが，近年では大規模なシミュレーションや機械学習，人工知能に関する手法が開発されてきた．パラメータ推定においても，潜在クラスモデルや階層ベイズモデルといった，計算コストの高い方法が広く使われるようになってきた．少しでも複雑な計算をしようとすると手元の PC では実行不可能であった時代からすれば格段の変化である．また，計算機・分析を取り巻く環境について，R や Python といったデータ分析を得意とする言語と集合知による各種分析手法のパッケージ公開が進み，提案とともにすぐにそれを実行できる環境が整ってきた．また，GitHub のようなプログラムプロジェクトのホスティングサービスや Qiita のような Q&A サイトの登場で，新しい手法もいち早く実装したり試行することができるようになり，分析の実行は格段に省力化できるようになった．前世紀には，ほとんどすべての手法は論文を読みつつ自力で実装しなければならなかった．途中（おそらく第 6 回あたり）に一度データハンドリングや分析手法の実装技術が発表の質にも影響を与えるということで，学生ならではの斬新で面白いアイデアを分析できるようにと，データ分析のツールを作っている企業にお願いをして分析ツールの貸出しの仕組みを作ったりもした．

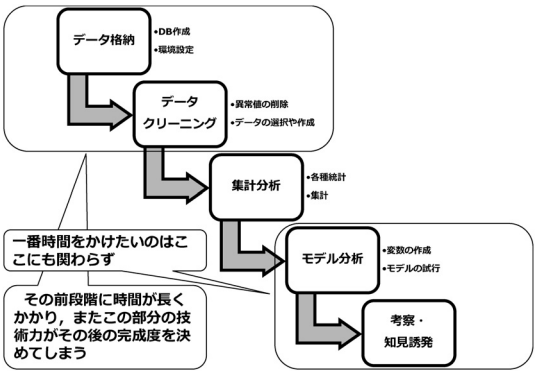

図 3　データ分析プロセス

データ分析の一般的なプロセスを図 3 に示すが，データ量が多くなると，分析そのものやその結果の吟味に費やす時間よりも前処理の時間に多くの時間が割かれることになる．また，そもそもそのような多くのデータをどのようにすれば使えるのかに途方に暮れるケースもあろう．分析ツールの提供はこのようなデータ量の増大に対する壁を乗り越えるための武器入手の方法になったかと思う．

このように，現在の参加者は昔と比べて，さまざまな武器をすでにもっており，これらを駆使できることも確かである．それに，学生は経験豊かな教員の指導の下で分析の目的ややり方を学びながら進めるため，すべてを一から始めるわけではないし，どうすれば分析を進められるのかは先生や先輩から教わることもできよう．場合によってはグリム童話の小人の靴屋に出てくる小人のように，教員がこっそりと（一番大変な）前処理をしているといったこともあろう．

ただし，「どういう変数を使うか（作るか）」「検証をどうするのか」「どのようなモデルを使うのか」ということについては，分析者が自ら決めなければならないし，また分析結果の解釈のためには，この解がなぜ得られているのかといったことを理解する必要があり，単にツールが使える，分析結果を求めることができるといったプログラミング能力だけでは十分とは言えない．

将来のデータサイエンティストを育成するために，教育する立場から考えると，さまざまな理論や技術の根本の理解と，それを効率的に駆使して分析を進めるというある種の二律背反な問題に対して，どのような形態でデータサイエンス教育をしていくのかは今一度考え直さなければならない部分もあるかもしれない．

5.　おわりに

25 年以上続けてきた本コンペティションについて，

本稿では特に後半において教育への影響について論じた．ICT の発展にも乗りデータサイエンスへの期待も変化し，それに合わせて関連する教育をどのように進めるべきかについては一層考えなければいけない時期にさしかかっているのかもしれない．くしくもコロナ禍において，大学の計算機環境も自由に使えず，またライセンス形態によっては学外からのリモートアクセスでは使えないツールもあると聞く．コンペティション活動のようにチームでデータ分析を進めるにおいては，これまでとは異なる工夫が必要とされる場も多いのではなかろうか？　本コンペティションもデータサイエンス教育のお手伝いをしてこれたかと思いつつも，ずっと続けているとある種の自家中毒に陥っているのではないかと思うこともあるため，今後の開催のあり方や期待について，ぜひ皆様からの叱咤激励をいただければ幸いである．本年度（令和 2 年度）も継続して行っているので，ご興味ある方はぜひとも発表会にご参加いただきたい．本学会では「データドリブンマーケティング研究部会」（主査：横山暁先生（青山学院大学），幹事：朝日弓未先生（東京理科大学），大竹恒平先生（東海大学））に開催をお願いしている．

謝辞　経営科学系研究部会連合協議会のメンバとして各研究部会を束ねていただいている先生方には，常日頃からのコンペティションの運営に大変なご協力をいただいております．また，毎回のデータを提供いただいてきた企業にも感謝申し上げます．ツール提供をいただいた各社，とりわけ（株）NTT データ数理システムには本当に長期にわたって毎年適切なツールを提供いただいております．最後に，私事ではありますが本学会よりコンペティション活動に対して小生に普及賞を授与いただきました．これは，本来は長年関係いただいている主催者一同が受け取るべきものと思いますが，今回代表して受け取ったと考えています．マーケティング，データ解析そして本コンペティションに関して深い理解をいただいている本学会へ感謝を申し上げ，本稿を閉じたいと思います．

参考文献

[1] 特集「スキャンパネルデータを用いたシェア予測」，オペレーションズ・リサーチ：経営の科学, Vol. 40, No. 9, 1995.
[2] 経営科学系研究部会連合協議会ウェブサイト，https://jasmac-j.jimdofree.com/（2020 年 7 月 31 日閲覧）
[3] kaggle ウェブサイト，https://www.kaggle.com/（2020 年 7 月 31 日閲覧）
[4] SIGNATE ウェブサイト，https://signate.jp/（2020 年 7 月 31 日閲覧）
[5] 特集「データ解析コンペティション」，オペレーションズ・リサーチ：経営の科学，Vol. 45, No. 12, 2000, Vol. 47–Vol. 65, No, 2, 2002–2020.
[6] 中田和秀，“データ解析コンペティションへの挑戦，”オペレーションズ・リサーチ：経営の科学，**63**, pp. 274–277, 2018.
[7] 須江雅彦，“我が国の未来を担うデータサイエンティストの育成―政策の動向と滋賀大学の挑戦―，”日本ソーシャルデータサイエンス学会論文誌，**1**, pp. 3–8, 2017.
[8] 文部科学省専門教育課，「数理・データサイエンス・AI 教育の全国展開」の協力校の選定について，2020. https://www.mext.go.jp/content/20200330-mxt_senmon01-000006307_1.pdf（2020 年 7 月 31 日閲覧）
[9] 鳥山明，「ドラゴンボール」，Vol. 1–Vol. 42，集英社，1985–1995.
[10] ドラクエパラダイスウェブサイト，http://www.dragonquest.jp/（2020 年 7 月 31 日閲覧）