今回の特集記事についてのコメント

生田目　崇

1. はじめに

巻頭の「特集にあたって」ですでに述べたように，今回の特集では平成23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論文募集を行い，9篇の投稿を得たものの，第1回目の査読で5編が返戻の判定となり，再査読判定4篇のうち1編は未投稿，残る3篇も残念ながら期日までに採録には至らなかった．これを受けて，投稿はなかったものの，課題フリー部門，課題設定部門で最優秀賞を受賞した2チームと，投稿論文のうちで第1回目の査読で再査読となり，再投稿のあった3編について，特集記事として掲載させていただいた．ここでは，それぞれの特集記事についてのコメントと，査読で指摘されたポイントを整理する．そして，近年の論文の投稿および査読の傾向などについて触れる．

2. 各特集記事について

2.1 松本氏・西郷氏の記事について

本稿を寄稿いただいた松本氏，西郷氏は数年来の本コンペティションに参加いただいている，いわば常連である．毎年驚かされるのは，業務外で本コンペティションに参加されていると聞いているが，毎回，質・量的に大変緻密な分析をされている点である．これまでは，コンペティションでの口頭発表とその配布資料でのみでその内容を知ってきたが，分析プロセスの詳細をまとめた文章は実は今回が初めてである．

なお，課題設定部門についての詳細については，「特集にあたって」で述べたので省略するが，スコアの提出のあった17チームから検証用データによる各会員セグメント予測値が提出され，採点した結果以下のようなスコア一覧であった．検証用データに含まれる会員数が16,563人であることから，上位5チームの予測差は100人から数百人程度であった．1位と2位の差はスコアでは0.01%であるが，正答数では200名程度の差があり，2位のチームはウェイトの高いセグメントの正答率が良かったものの総合成績ではわずかに及ばなかった．逆にいえば，松本氏らの記事にもあるように「バランスよく」予測できるようにモデル構築をしたことが，スコアに反映されている．今回記事を読ませていただいて大変感銘を受けたのは，分析デザインについて，かなり多面的に計画を立てていることである．多くのチームが割と事前に断定的に分析を行っていると聞いているなかで，あらゆる可能性を考えて，よりよいモデルを探索していることは興味深い．日頃から実データを元に分析されているとのことで，このあたりの勘所はもたれているのかもしれないが，実際にこれだけの緻密な分析をされたことには脱帽せざるを得ない．特に最後に予測精度が上がらないセグメントのみ追加的にモデルを変えているところなどは興味深い．なるほどここまでやらないと太刀打ちできないのか，と自分自身主催者ではあるものの大変感心させられた．

表1　課題設定部門獲得スコア一覧

順位	スコア率	順位	スコア率
1	70.43%	10	64.03%
2	70.42%	11	63.89%
3	69.43%	12	59.50%
4	69.33%	13	58.24%
5	68.36%	14	47.49%
6	65.72%	15	42.33%
7	65.53%	16	36.76%
8	64.65%	17	21.40%
9	64.37%		

2.2 鮭川氏らの寄稿について

鮭川氏らのチームも，論文の投稿はなかったが，大変興味深い成果であるため寄稿をお願いした．鮭川氏らのグループも，筑波大学の山本芳嗣教授門下の古くから参加のあるチームであるが，学生主体のチームであるため，メンバーは毎年変わる．初代メンバーの高野祐一氏（現・東京工業大学）にお聞きしたところ，毎週定例ミーティングを行って進捗を確認しつつ，次の課題を出しながら進めているということで，やはり大

なまため　たかし
専修大学商学部
〒214–8580　神奈川県川崎市多摩区東三田2–1–1

変熱心に分析にあたっている．

今回の発表では，ほぼすべてのチームが，データをいかに分析するかに主眼を置いているなかで，分析をするためツールを作成したということが成果報告会当日は相当驚かれた．データ提供元の（株）ゴルフダイジェスト・オンライン社からも当日は審査に参加いただいたが，顧客行動の可視化ツールとして大変興味深いという評価をいただいた．本コンペティションに長年携わっていると，どうしても連作障害的に同じような視点での分析をしがちななかで，こうした新たな視点の発表があることはわれわれにとっても大変刺激的である．何より驚かされたのは，こうしたシステムのプロトタイプをわずか1週間で作りあげたと聞いたことであり，プログラミング技術についてはもうすでに置いて行かれた…とつい思ってしまったことである．

このツールについてもただデータを流し込んで様子を見るだけではなく，さまざまなレベルの試行錯誤ができるようになっており，今回得られた3つの知見などは，まさに，「見える化」による賜物と言えよう．

2.3 佐藤氏，朝日氏の記事について

本稿は，ウェブサイトへの来訪行動とそのうえでの購買行動の2段階の行動を確率選択モデルとして表現し，実データを当てはめた解析を行っている．そして，個人差を考慮するために，サンプルごとのパラメータを求める，階層ベイズモデルを用いている．得られた結果についてROC曲線によってモデルの有効性を確認している．

本稿について査読者からは，説明変数の妥当性，提案モデルの有効性についての指摘があった説明変数については，本稿で採用した説明変数が主観的に選ばれているという疑念が払拭できないという指摘と，モデルの有効性については，既存モデルと比較の結果，全体の識別能力が向上していないことは，モデル構築が適切でない可能性があるという指摘がなされている．

顧客行動の記述としては，大変興味深いモデリングであるとはいえ，査読者の指摘のとおり，モデルそのものを既存の論文を踏襲したものであり，もう一歩踏み込んだ議論ができなかったことは残念である．後述するように，現在のマーケティング・サイエンスの論文の多くが個人差に着目したものであり，本稿は消費者行動分析分野の先端的なアプローチの一つであるとは言えよう．

2.4 佐賀氏らの記事について

本稿では，逐次確率比検定によるトレンド変化点の検出と，クラス遷移と共起遷移の両分析を含んだFACT-Graphを用いて，商品のカテゴリをより効果的に把握して分析するために階層的FACT-Graphを提案している．日々のサイトアクセス数の推移をデータとし，提案した手法をあてはめることによって，13の変化点を見出し，その変化点で区切られた区間ごとのカテゴリ購買に対して一定知見を得ている．このように統計的な判断とグラフによる購買状況の可視化によって，商品の注目度の傾向や共に考慮されている商品などの把握が可能になるという貢献が確認できる．

しかし一方で，統計的検定の扱いについての疑問が査読者から指摘された．具体的には，確率変動するとするデータに対して回帰直線を当てはめると，回帰直線および分散も確率変動するはずであるが，そのことは検定に含まれていないこと，対象データは連続変数を仮定しているにもかかわらず，提案手法では離散確率を元にした判断を行っており，論理的な整合性が確認できないというものであった．これらの指摘はあるものの，野心的な研究でありアクセスログ・データ分析に新たな一石を投じる可能性もあったことから，採録にまで至らなかったことは大変惜しかった．

2.5 久松氏らの記事について

本稿は，どういった訪問時に購買は発生するか，また消費者が具体的に購買に向かっているのかについて判別しようというモデルを提案している．ただし，会員の訪問行動はさまざまであるため，個人差を考慮するために潜在クラスモデルを用いたセグメンテーションをおこなっている．そのうえで，各訪問時における行動変数による，購買生起の確率モデルを提案しその評価を行っている．その際に所属確率をウェイトとして尤度を求めている．

確率モデルの尤度計算やモデル選択などにおいては一定の評価がされたものの，査読で指摘された一番のポイントは学習用データと検証用データの分割方法とその評価であった．投稿論文においては，1年間のデータを会員ごとに分割し，片方を学習用，もう片方を検証用としてモデル構築と評価が行われていたが，この分割方法では新たな会員に対して評価をすることができないという点に対する指摘に対して有効な回答が得られなかった．また，今回のデータが一年であることから購買の季節変動などへの対応についての不足も指摘されている．

3. おわりに

今回，9篇の投稿を得たがこれはここ数年から見ても平均的な数であり，特別に少ないというわけではな

かった．ただし，第一回目の査読結果で5篇と，半数以上の論文に返戻という判定になったことは，例年にはなく厳しい判定となった．

今回の特集における査読については，ダブルブラインド方式で公平性を保つことはしているものの，査読期間が短く，また分野も限定されていることから，事情を理解いただいてご協力いただける査読者に依頼がどうしても偏ってしまいがちである．したがって過去の本特集の査読をされた方の場合は，マーケティングのデータ解析分野に対する慣れによって査読が厳しくなったり，前に査読した論文のレベルが頭をよぎったりすることもあるかもしれない．

また，投稿者についても半数は過去に投稿経験のある方々であり，投稿にチャレンジするという意気込みは評価できるものの，反面，毎年参加しているなかでの悪い意味での慣れが出てきている可能性も考えられる．このことは，投稿のみならず発表を聞いていても，いわば既視感を感じることもあり，こうした成果を元にした論文に対しては，自然と査読が厳しくなりがちになることも十分に考えられる．

ただし，いずれにしても，厳しく査読をしていただくことは，本特集の質を保持するうえでは必要不可欠なステップであり，逆にこうした厳しい査読を通る論文にこそ価値があるものと考えられる．今後いつまでこうした特集が組めるかはわからないが，続く限りは論文のレベルを維持もしくはアップできるような体制を続けていきたい．

今回，投稿論文から特集記事に回ってもらった3篇の寄稿においては，査読の指摘に対してもう一歩踏み込んだ分析や考察もしくはモデリングがなされていれば十分に採録に値したと思われる．逆に，一気通貫した論理性や，データの適切なハンドリングを経ないと採録に至らないということである．本コンペティションは，企業のリアルなデータを提供していただくが，参加者は広く募集するため（ただし，参加にあたってはコンペティション事務局およびデータ提供元で身元のチェックは行う），どうしてもマスキングする項目がでたり，母集団に対して歪んだサンプリングをしたりする場合がある．こうした，限定された情報をいかに駆使して知見を得るか，またビジネス課題を設定しどのようにそれをデータから解決するかについて，具体的かつ汎用可能性のあるモデルづくりや分析が必要となる．「アイデアはいいけど，ちょっと何かが不足している」というような場合，間違いなく査読者が見逃さず，それを指摘されている．

近年では，マーケティングのモデル分析はより複雑になっていると言ってもよいであろう．原因はさまざまにある．ビッグデータに代表されるように，ある観測対象について，測定データ元の異なるような高次元で粒度の細かいデータが取得できるようになり，それを分析できるような計算機環境や分析技術が発達してきたことも挙げられる．またその反面，そうしたデータについて粒度を落とすことなく使わなければならないといった風潮もあるのかもしれない．コンペティションの発表や投稿論文においても，複雑なモデルや個人別パラメータを求めるといったものが増えてきているが，当初こうした発表を聞いたときの驚きを今になっても思い出す．

日本におけるマーケティング・サイエンスを引っ張ってきた，中西正雄関西学院大学名誉教授によれば，アメリカの学会では，個人パラメータを求める階層ベイズを使わなかったら論文にならないといった風潮があるが，実際にそういったプログラムをみんなが書いているわけではない，といったニュアンスの指摘もされている [1]．つまり，分析環境が進化する反面，一から十まで分析者が責任を持ってプログラミングするという環境ではなくなり，ある種ブラックボックス化されているということに対する警鐘である．実際に自ら手を動かして分析し理解するといった姿勢が重要であるにもかかわらず，いわば，計算機環境の中に翻弄されている様子というのがここには示されているとも言える．私自身も，こうした手法の恩恵にあずかることは少なくないものの，得られた解が最適解であるのかについて疑念をもつこともしばしばである．

たしかに構造的に複雑なモデルを手軽に解くことができるような計算機環境が整備されたことは，データ解析技術の大きな進展であることは間違いないが，果たしてその環境の中だけにとどまってしまっていいのか，もしかしたら今再び考えるべき時点にきているのかもしれない．

最後になったが，今回のデータ解析コンペティションにおいて，データを提供いただいた，（株）ゴルフダイジェスト・オンライン，分析ツールを提供いただいた，（株）数理システム両社には多大なご協力に感謝申し上げたい．

参考文献

[1] 中西正雄，川上智子，石淵順也（対談），「データをマッサージする」，碩学舎ビジネスジャーナル，碩学舎，**6**，2012.