

目指せデータサイエンティスト

— 学生が競うデータ解析コンペティション —

東海大学 理学部 数学科 教授 やまもと **山本** よしろう **義郎**

平成29年5月11日（木）、12日（金）に、東京理科大学森戸記念館で開催された日本計算機統計学会第31回大会において、特別セッション「データカフェ」で3件のデータ解析事例についての報告が行われました。

このデータカフェセッションは、経営科学系研究部会連合協議会が主催する「データ解析コンペティション」に、日本計算機統計学会スタディグループ部門として出場したチームからのデータ解析事例が報告され、今回は、明治学院大学心理学部のチームによる「4WによるECサイト利用者の購買行動分析」、東海大学理学部数学科のチームによる「アンケート回答傾向による優良顧客へのアプローチ」、徳島文理大学理工工学部のチームによる「アンケートと購買金額による会員のクラスタリング」の3件が報告されました。

経営科学系研究部会連合協議会主催の「データ解析コンペティション」は、平成6年より「共通の実データを元に、参加者が分析を競う」ことを目的として開催され、POSデータのような顧客の購入データや、生活者の意識を調査したアンケートデータなどが提供され、20年以上毎年開催されています。

このコンペティションは、当初は学会研究部会単独で開催したが、参加する研究部会が増えたため、現在は経営科学系研究部会連合協議会という研究部会の連合により運営され、構成研究部会単独もしくは合同で予選を実施し、各部会から選抜されたチームが全体の成果報告会で発表できる、全体としての最

優秀賞や優秀賞が選ばれます。「データ解析コンペティション」のWebサイト (<https://jasmac-j.jimdo.com/>) では、過去の年度の分析データの種類についての情報と全体としての入賞グループについての情報を確認することができますので、興味をもたれましたら是非調べてみて下さい。

ちなみに、前回の平成28年度のコンペティションはECサイトの購買データと一部の顧客に対するアンケートデータが提供され、約100チームが参加し、各部会の予選を勝ち抜いた13チームが平成29年3月10日に開催された最終報告会に進出し、報告を行いました。日本計算機統計学会スタディグループは16チームが参加し、その中から2チームが最終報告会に進出しました。写真は、平成29年3月2日に統計数理研究所で開催された、日本計算機統計学会スタディグループ部門の最終報告会での発表の様子です。

データ解析コンペティションが始められた二十数年前は、1990年頃から始まったコンピュータの急速な性能向上とインターネットの普及から、データ分析に対する期待が高まり、「データマイニング」の重要性に注目が集まっていました。実際のデータを分析して、経験を積みたいという要望はあるものの、実際にデータマイニングをする機会は企業での分析担当者だけという状況でした。そのような状況に対応するために、データ解析コンペティションが始まりました。このコンペティションでは、学生のグループ（多くは



写真 平成28年度データ解析コンペティション 日本計算機統計学会スタディグループ部門の報告会

大学院生)と社会人大学院生も含む社会人の参加があったため、学生部門と一般部門とに分かれて入賞に関する審査を行っていました。2年ほど前から学生部門と社会人部門との差があまりなくなったため、学生も社会人も一緒に審査されることになりました。

日本計算機統計学会も、平成21年度からデータ解析コンペティションに参加し、最初は参加チームも少なかったのですが、最近では参加人数の多い部会の1つになっています。また、他の部会では学生の参加は大学院生が多いのですが、日本計算機統計学会スタディグループ部門は、学部学生それも学部3年生の参加も多いのも1つの特徴となっています。3年生でコンペティションに参加すると、4年生での就職活動に大変有利となり、また、もっと高度な分析を理解し、自分で新たな分析手法を開発するために、大学院に進学するきっかけになる学生も多いです。

また、参加大学の地域性も全国規模で、関東の大学だけでなく、北海道大学、岡山理科大学、福岡女子大学、九州大学などが参加しているため、いろいろな大学の学生と交流できるという特徴もあります。同じデータを分析しているので、お互いに苦労した点などを話すと話は尽きないといったところでしょう

か。さらに、各大学の学部学科も、理学部数学科、数理情報科学科など数学や情報の分野ばかりではなく、国際文化学部などの学際系(文理融合)の学部や、心理学部などの文化系の学部もあり、さまざまな学部でデータサイエンスに興味をもって学んでいる様子もうかがうことができます。

中学校の数学では「資料の活用」、高等学校では「データの分析」において、データを分析するための授業が行われていますが、多くの中学生、高校生にとっては、データ分析は必ずしも身近なものでないかと思います。現在は、一部のデータ分析を行う人達がデータサイエンティストと呼ばれて、データ分析能力を必要とされていますが、今の高校生が社会人となる頃には、多くの人が統計科学の基礎的な考え方(統計リテラシー)、コンピュータを自由に使いこなす、さまざまな分野での問題解決に取り組む時代になっているような気がします。

データサイエンティストを目指す学生には、このような面白い挑戦の舞台が増えています。中高生でも参加できるコンペティションなども増えていますので、自分が参加できるものがあったら是非挑戦してみてください。